



Contribution informatique en sciences pour l'environnement

Christophe Paoli

► To cite this version:

Christophe Paoli. Contribution informatique en sciences pour l'environnement. Réseau de neurones [cs.NE]. Université de Corse, 2014. tel-01288151

HAL Id: tel-01288151

<https://hal.science/tel-01288151>

Submitted on 14 Mar 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Public Domain



Université de Corse - Pasquale Paoli
Université Galatasaray



UMR CNRS 6134
Sciences Pour l'Environnement

Ministère des Affaires Etrangères



Synthèse de l'activité scientifique en vue d'obtention de l'Habilitation à Diriger des Recherches

Soutenue publiquement par :
Dr. Christophe PAOLI
le 01 septembre 2014 à l'Université de Corse

Contribution informatique en sciences pour l'environnement

Directeur :

Dr. Jean-François MUZY, Directeur de Recherche au CNRS

Rapporteurs :

Prof. Philippe LAURET, Université de La Réunion

Dr. HDR Fawaz MASSOUH, Arts et Métiers ParisTech

Dr. Jean-François MUZY, Directeur de Recherche au CNRS

Jury :

Prof. Marc MUSELLI, Université de Corse,

Prof. Philippe LAURET, Université de La Réunion

Dr. HDR Fawaz MASSOUH, Arts et Métiers ParisTech

Dr. Jean-François MUZY, Directeur de Recherche au CNRS

Dr. HDR Dominique LAMBERT, Maître de Conférences à l'Université de Toulouse

Dr. HDR. Gilles NOTTON, Université de Corse.

Président :

Prof. Marc MUSELLI, Université de Corse

Table des matières

I. Remerciements	8
II. CV détaillé	9
A. Etat Civil	9
B. Etudes secondaires et universitaires	9
C. Déroulement de carrière	10
D. Synthèse des activités de recherche	10
E. Synthèse des activités administratives et pédagogiques	13
F. Synthèse des activités d'enseignements	15
G. Activités de consultance	16
H. Autres activités	16
III. Résumé des travaux de recherche et d'encadrement	19
A. Activités en matière de recherche	19
B. Points forts des activités de recherche	20
C. Valorisation : contrats, dépôts de brevets, logiciels	23
D. Information scientifique, technique et vulgarisation	23
E. Activités internationales : conférences invitées, contrats, séjours à l'étranger, accueil de collègues et stagiaires étrangers	24
IV. Charges d'enseignement occupées	26
A. Université Galatasaray (GSÜ)	27
B. Université de Corse - Faculté des Sciences et Techniques (FST)	27
C. Université de Corse - Ecole d'Ingénieur « Paoli Tech »	29
D. Université de Corse - Formation Continue	29
E. Rectorat de Corse	29
F. Centre départemental de gestion de la fonction publique territoriale	30
V. Fonctions d'intérêt collectif exercées	31
A. Responsabilités scientifiques	31
1. Encadrements de thèses	31
2. Participation à des encadrements de thèses	31
3. Encadrements de master	32
4. Participation à des encadrements de master	33
5. Participation à des projets scientifiques	33

6.	Référés et conférences internationales	34
7.	Fonctions électives	35
8.	Divers	35
B.	Responsabilités administratives	35
1.	Directions	35
2.	Fonctions pédagogiques	36
3.	Fonctions électives	37
VI.	Liste des publications	38
A.	Publications dans des revues internationales à comité de lecture de Rang A	38
B.	Publications dans d'autres revues internationales à comité de lecture	39
C.	Publications dans des revues nationales à comité de lecture	39
D.	Participation à des ouvrages collectifs	39
E.	Communications avec actes	40
F.	Rapports divers	43
VII.	Activités de recherche	44
A.	Avant-propos	44
B.	Décidabilité et calculabilité	46
C.	Prédiction et modélisation de phénomènes spatio-temporels	47
1.	Introduction	47
2.	Les modèles de prédiction	49
a)	Les modèles de prévision numérique du temps	49
b)	Les modèles statistiques	51
c)	Comparaison des performances	54
d)	Différentes méthodes pour différentes échelles	55
3.	Problématiques relatives aux méthodes statistiques	57
a)	Prétraitement des données	57
b)	Sélection de caractéristiques	58
c)	Hybridation et sélection de modèles	60
d)	Synthèse	61
4.	Applications au rayonnement solaire et aux pics de pollution	62
a)	Prédiction de séries temporelles de rayonnement solaire global et de production d'énergie photovoltaïque à partir de réseaux de neurones artificiels	62
b)	Prévision des concentrations de polluants atmosphériques en Corse et impacts sur la productivité des systèmes énergétiques solaires	68
c)	Utilisation des réseaux de neurones artificiels pour l'estimation du potentiel solaire	74
5.	Synthèse	78
D.	Aide à la décision	79
1.	Concepts de base de l'Aide MultiCritère à la Décision	79
2.	Etat de l'art et constats	81
3.	Concepts de base du web sémantique	81
4.	Représentation de la connaissance avec les langages RDF-RDFS, OWL et SKOS	83
5.	Synthèse et perspectives	83
6.	Conclusion	85

E. Test de logiciel	86
1. Problématique	86
2. Description de l'approche	87
a) Exploration des techniques de test de logiciels	88
b) Adaptation de ces techniques à VHDL	88
c) Implémentation du prototype logiciel GENESI	90
3. Synthèse et conclusion	91
VIII. Conclusion générale	92
A. Bilan et conclusions	92
B. Perspectives de recherche	93
IX. Références	95

Table des illustrations

Figure 1. Exemple de maillage pour la région des Alpes pour des maillages différents.	50
Figure 2. Neurone formel et architecture d'un PMC.	52
Figure 3. Construction d'un PMC utilisable dans le cas de prédiction de séries temporelles.	53
Figure 4. Classification des méthodes de prédiction en fonction de la résolution spatiale et temporelle.	55
Figure 5. Procédure générale de sélection d'un sous-ensemble de caractéristiques.	59
Figure 6. Vue schématique de la formalisation du problème de la sélection d'algorithmes.	60
Figure 7. Localisation des 5 villes étudiées.	62
Figure 8. Principe de la méthodologie hybride.	65
Figure 9. Situation géographique des stations de surveillance en Corse (fournie par l'IGN).	69
Figure 10. Concentration d'ozone observée (en bleue) et prévue 24h à l'avance (en rouge) en aout 2013 à la station Giraud (Bastia).	70
Figure 11. Concentration d'ozone observée (en bleue) et prévue 24h à l'avance (en rouge) en aout 2013 à la station Canetto (Ajaccio).	71
Figure 12. Principe des modèles prévisionnels avec classification.	72
Figure 13. Radiation incidente sur une surface horizontale et inclinée.	75
Figure 14. Différence entre le web actuel et le web sémantique.	82
Figure 15. Processus de conception d'un circuit complexe.	86
Figure 16. Aperçu de notre approche.	87

Table des tableaux

Tableau 1. Caractéristiques principales des données de rayonnement global horizontal. _____	63
Tableau 2. Synthèse des résultats concernant le rayonnement solaire global et de production d'énergie photovoltaïque. _____	64
Tableau 3. Descriptif des stations de mesure en Corse. _____	69
Tableau 4. Performances de la prévision à J + 1 des modèles basés sur les PMC et du modèle AIRES, pour l'O ₃ et les PM ₁₀ . _____	71
Tableau 5. Caractéristiques principales des données disponibles à Vignola et Bouzareah. _____	76
Tableau 6. Structure de la matrice de performance. _____	80



Université de Corse - Pasquale Paoli
Université Galatasaray
-
UMR CNRS 6134
Sciences Pour l'Environnement
Ministère des Affaires Etrangères



"Prediction is very difficult, especially if it's about the future".

Niels Bohr

"Everything should be made as simple as possible, but not simpler".

Albert Einstein

I. Remerciements

Je tiens tout d'abord à remercier Philippe LAURET, Fawaz MASSOUH et Jean-François MUZY d'avoir accepté de relire ce mémoire et d'en être rapporteurs. La version finale de ce mémoire a bénéficié en outre de la lecture particulièrement attentive et des remarques nombreuses et précieuses de Cyril, Gilles, Marie-Laure, Valérie et Wani. Je remercie également tous les autres membres du jury : Dominique LAMBERT, Marc MUSELLI et Gilles NOTTON d'avoir accepté d'assister à la présentation de ce travail. Merci aussi à Marie-Christine, Marie-Laure et Philippe pour avoir pu assurer la logistique.

Les travaux présentés dans ce mémoire ont été réalisés essentiellement au Laboratoire Sciences Pour l'Environnement (SPE) de l'Université de Corse, dirigé par Jacques-Henri BALBI puis par Paul-Antoine BISGAMBIGLIA. Je les remercie de m'avoir accueilli au sein du SPE et de m'avoir permis de m'exprimer dans le cadre passionnant du métier d'enseignant-chercheur.

Détaché à l'Université Galatasaray (Istanbul, Turquie) depuis septembre 2013, j'ai pu bénéficier de la compréhension de ma faculté d'accueil et des services du Ministère des Affaires Etrangères et du Développement International pour dégager le temps nécessaire à la rédaction de ce mémoire. Je remercie ainsi chaleureusement l'ensemble de mes collègues de l'Université Galatasaray et du Ministère pour leur compréhension et la liberté qu'ils ont su m'accorder tout au long de cette année.

Je tiens également à remercier tous les étudiants, doctorants, stagiaires et alternants que j'ai pu encadrer ou co-encadrer tout au long de mon parcours universitaire. Leur travail, leur implication, leur force et leur motivation ont largement contribué aux résultats décrits dans ce rapport.

Enfin, je tiens à exprimer mes remerciements à ma famille et mes amis pour leur soutien et affection.

A tous merci !

II. CV détaillé

Christophe PAOLI : Docteur - Maître de Conférences classe normale - Informatique -
Section 27 - Université Galatasaray - Faculté d'Ingénierie et de Technologie -
Département de Génie Informatique.

A. *Etat Civil*

Né le 26 septembre 1973 à Bastia (20200)
Situation familiale : pacsé

Nationalité : Française
3 enfants

Adresse au laboratoire

Université de Corse
UMR CNRS 6134
Campus Grimaldi
Bât. Alfonsi, 20250 Corte
☎ : 04 95 45 02 09
✉ : christophe.paoli@univ-corse.fr
🌐 : <http://spe.univ-corse.fr>

Adresse administrative

Université Galatasaray
Ciragan Cad. No:36,
Ortaköy 34349 Istanbul
☎ : +90 (212) 227 44 80
✉ : cpaoli@gsu.edu.tr
🌐 : <http://gsu.edu.tr>
🐦 : twitter.com/cpaoli

B. *Etudes secondaires et universitaires*

- **Doctorat** en informatique, intitulé « Validation de descriptions VHDL fondée sur des techniques issues du domaine du test de logiciels », obtenu le 20 décembre 2001 à l'Université de Corse avec mention très honorable devant un jury composé de :
 - Prof. Matteo SONZA REORDA, Politecnico di Torino, Président ;
 - Prof. Denis FLOUTIER, École Supérieure d'Ingénieurs d'Annecy, rapporteur ;
 - Prof. Andrzej RUCINSKI, University of New Hampshire, rapporteur ;
 - M. Paul BISGAMBIGLIA, Maître de Conférences (HDR) à l'Université de Corse ;
 - Mlle Marie-Laure NIVET, Maître de Conférences à l'Université de Corse ;
 - Prof. Jean-François SANTUCCI, Université de Corse, Directeur de thèse.
- **DEA** « Mécanique Énergétique » - 1998 - Mention Bien.
- **Licence et Maîtrise** « Physique et Applications » - 1995/97 - Université de Corse.
- **DEUG A** « Sciences et Structures de la Matière » - 1994, Université de Corse.
- **BTS** « Assistant Ingénieur » - 1993 - Lycée Paul Vincensini de Bastia.
- **Baccalauréat** - Série D - 1991 - Lycée polyvalent de Porto-Vecchio.

C. *Déroulement de carrière*

- **Depuis 2013** : Nommé Expert Technique International (ETI) par le Ministère des Affaires Etrangères et du Développement International à l'université Galatasaray (Turquie) en tant qu'enseignant chercheur en informatique au sein du département de Génie Informatique de Faculté d'Ingénierie et de Technologie.
- **2013** : Obtention d'un accueil en délégation au CNRS (section 27) - UMR CNRS 6134.
- **2009-2013** : Directeur du département informatique de la Faculté des Sciences et Techniques (FST) de l'Université de Corse.
- **2010** : Obtention d'un 1/2 CRCT par voie CNU (section 27).
- **2002-2008** : Vice doyen de la FST en charge de l'insertion professionnelle.
- **2005-2006** : Responsable de la mention informatique de la licence sciences et technologies à la FST de l'Université de Corse.
- **2002** : Maître de Conférences à la FST.
- **2001-2002** : Contrats de 1/2 ATER puis d'ATER complet en informatique à la FST de l'Université de Corse.
- **2002** : Qualification aux fonctions de Maître de Conférences (section 61).
- **2001** : Obtention d'un doctorat en informatique (Université de Corse).
- **1998-2001** : Doctorant boursier et moniteur (formation généraliste au métier d'enseignant chercheur sous la forme de stages) en informatique à l'Université de Corse.

D. *Synthèse des activités de recherche*

(Pour plus de détails cf. sections III "Résumé des travaux de recherche et d'encadrement" et VII "Activités de recherche")

Depuis 2005 : Membre du projet « *Energie Renouvelable* » de l'UMR CNRS 6134 - Université de Corse : utilisation de techniques d'intelligence artificielle pour la modélisation et la prédiction de séries temporelles, représentation des connaissances et aide à la décision.

- **Encadrement d'étudiants en Doctorat** :
 - Co-direction de M. Frédéric GUIDANA GAZAWA : « Evaluation multicritère et e-gouvernance territoriale des ressources naturelles au Cameroun », démarrée en juin 2012.
 - Co-direction de M. Wani TAMAS : « Prévion des concentrations de polluants atmosphériques en Corse et impacts sur la productivité des systèmes énergétiques solaires », démarrée en mai 2012.
 - Co-direction de M. Cyril VOYANT : « Prédiction de séries temporelles de rayonnement solaire global et de production d'énergie photovoltaïque à partir de réseaux de neurones artificiels », soutenue le 16 novembre 2011, qualifié CNU 62.

- Participation aux travaux de Mlle Kahina DAHMANI : « Utilisation des réseaux de neurones artificiels pour l'estimation du potentiel solaire en Algérie », thèse démarrée en 2012.
- Participation aux travaux de M. Ionut CALUIANU : « Augmentation de la productivité énergétique des modules photovoltaïques », soutenue le 06 Septembre 2011, mention exceptionnelle.
- **Encadrement d'étudiants en Master**
 - Responsable scientifique du contrat d'apprentissage de Mlle Aurelia BALU en master informatique : « Techniques de l'intelligence artificielle et développement logiciel dans le cadre de la prédiction de séries temporelles », démarré en septembre 2012 pour une durée de 2 ans.
 - Encadrement de M. J-B. MARCHETTI : « Mise en œuvre d'une application informatique visant le développement et l'aide à la localisation multicritère de projets photovoltaïques en Corse », Université de Corse, stage de 5 mois, 2009.
 - Encadrement de Mlle S. GANCHEVA : « Comparaison entre différentes méthodes pour la prédiction de vitesses de vent : Réseaux de Neurones Artificiels et modèles ARMA », Université Technique de Sofia, stage de 5 mois, 2007.
 - Encadrement de Mlle V. ANDRONOVA : « Utilisation de données météo et des réseaux de neurones pour la prédiction de vitesses de vent », Université Technique de Sofia, stage de 5 mois, 2006.
 - Encadrement de M. M. MAZOUZ : « Développement d'un plugin logiciel et intégration au sein d'une plate-forme générique d'aide à la décision », Université de Corse, stage de 5 mois, 2012.
 - Co-encadrement de M. W ZHAO : « Prediction of wind energy capacity using artificial intelligence methods », école des Arts et Métiers ParisTech, stage de 5 mois, 2013.
 - Co-encadrement de M. N.A. KLIMOV : « Utilisation des réseaux neuronaux pour l'estimation des données d'irradiation solaire directe », Université Technique de Sofia, stage de 5 mois, 2012.
 - Co-encadrement de Mlle L. IVANOVA : « Utilisation des réseaux neuronaux pour l'estimation des données d'irradiation solaire sur des plans inclinés », Université Technique de Sofia, stage de 5 mois, 2011.
 - Co-encadrement de Mlle S. VASILEV : « Application des réseaux neuronaux à l'estimation du rayonnement diffus », Université Technique de Sofia, stage de 5 mois, 2011.
 - Co-encadrement de M. M. PADOVANI : « Mise en place d'un prédicteur de pics de pollutions atmosphériques à partir de réseaux neuronaux », ENSAM ParisTech, Aix en Provence, stage de 5 mois, 2010.
 - Participation aux travaux de Mlle M. BATTISTINI : « Etude statistique pour l'élaboration d'un modèle de prévision de l'indice de la qualité de l'air en Corse », Ecole des Mines de Douai, stage de 5 mois, 2010.

- Participation aux travaux de Mlle P. RANDIMBIVOLOLONA : Prédiction du rayonnement solaire global horizontal à l'horizon H+24 », Mines ParisTech, stage de 6 mois, 2010.
- Participation aux travaux de M. M. UGER : « Intérêt des techniques de stockage pour le développement des ENR à caractère aléatoire », Mines ParisTech, stage de 6 mois, 2009.
- **Participation à des projets de recherche**
 - Participation au programme Hubert Curien (PHC) PLATON 2013, Projet 30355UE, Développement de modèles durables d'approvisionnement en énergie et en eau pour les îles méditerranéennes par utilisation des EnR.
 - Chercheur au sein de la plateforme MYRTE (Mission hYdrogène Renouvelable pour l'inTégration au réseau Electrique), ~21M € : prévision à courte et très courte échéance de l'énergie produite dans le but d'optimiser la gestion du stockage hydrogène dans les réservoirs et de piloter au plus fin le système énergétique en fonction de la charge à alimenter ; co-animateur du site web dédié : <http://myrte.univ-corse.fr>.
 - Chercheur et responsable scientifique au sein de l'observatoire CORSICA (« Centre d'Observation Régional pour la Surveillance du Climat et de l'environnement Atmosphérique et océanographique en Méditerranée occidentale ») ~2,2 M € : prévision des concentrations de polluants atmosphériques en Corse.
 - Dépositaire et référent d'un contrat de collaboration de recherche entre le laboratoire SPE et l'association agréée pour la surveillance de la qualité de l'air, Qualitair Corse : étude sur le thème de la prévision de la qualité de l'air et élaboration d'un outil statistique et informatique de prévision.
 - Coordinateur local du chantier MISTRALS (« Mediterranean Integrated STudies at Regional And Local Scales »), de l'Alliance ALLENI et de son axe ChArMEx (« the Chemistry-Aerosol Mediterranean Experiment »), initiative de l'INSU.
- **Responsabilité dans des congrès, des revues ou groupements scientifiques**
 - Président du comité d'organisation et webmaster du 3^{ème} workshop international ChArMEx, 5-8 novembre 2012 à l'IESC de Cargese, <http://charmex.lsce.ipsl.fr>.
 - Président du comité d'organisation et membre du comité scientifique de la 8^{ème} édition du « Decision Deck Workshop » et de la 73^{ème} réunion du groupe de travail européen « Aide Multicritère à la Décision », webmaster du site web dédié : <http://mcda.univ-corse.fr>.
 - Référés pour la revue « Renewable Energy » (Elsevier, 5-Year-IF=3.2) : <http://ees.elsevier.com/rene>.
 - Référés pour la revue « Journal of the Air & Waste Management Association » (Taylor & Francis, IF=1.204) : <http://www.awma.org>.

- Membre du comité scientifique et référés pour la conférence IEEE « International Conference on Environment and Electrical Engineering » : <http://eeeic.eu>.
- Référencement de nos travaux en matière de prédiction au sein de l'action COST (« European Cooperation in Science and Technology ») WIRE (« Weather Intelligence for Renewable Energies ») rassemblant différents acteurs du domaine dans le but d'améliorer les techniques de prévision météorologique dédiées aux énergies solaires et éoliennes.

Jusqu'à l'année 2005 : *UMR CNRS 6134 - l'Université de Corse : Utilisation de techniques de test de logiciel pour valider des systèmes électroniques décrits en langage VHDL, « VHSIC Hardware Description Language ».*

- Responsable scientifique et coordinateur du contrat de recherche « Développement d'un outil de test de circuit et amélioration de la testabilité des circuits », financé par la CTC « Collectivité Territoriale de Corse », ~16k €, 2003-2004.
- Participation comme chercheur au contrat de recherche international N°: F61775-00-C0002, projet 00-4005, item 0003 : « Validation of VHDL description », financé par l'EOARD : « European Office of Air Force Research and Development », ~15k €, 2000-2003.
- Participation comme chercheur aux workshops international ATW (« Advance Technology Workshop ») organisés par l'Université de Corse et sponsorisé par la CTC et l'US Air force : conférences pluridisciplinaires proposant de mettre en synergie universitaires et industriels, 1998-2000.
- Programme d'échange international (stage de 3 mois) au sein du laboratoire « Electrical and Computer Engineering » de l'UNH, « University of New Hampshire » (Durham, NH, USA) sur le projet : « Overview of Test Methods Using Boundary Scan », 1999.

E. Synthèse des activités administratives et pédagogiques

(Pour plus de détails cf. section V : Fonctions d'intérêt collectif exercées)

Depuis 2013 : *Expert Technique International (ETI) à l'Université Galatasaray (GSÜ) en Turquie au sein département de Génie Informatique de Faculté d'Ingénierie et de Technologie.*

- Structuration, animation et développement des relations entre GSÜ et les acteurs de la relation économique franco-turque.
- Suivi de l'utilisation d'applications informatiques au niveau de l'université, notamment d'une plateforme d'enseignement en ligne de Français Langue Etrangère (FLE).
- Cartographie des activités de recherche de GSÜ à travers un annuaire.
- Aide à la conception et à la mise en œuvre de coopérations avec les établissements français membres du consortium Galatasaray.

- Coordination de 7 enseignants français intervenant au sein des départements de génie informatique et industriel de la Faculté d'Ingénierie et de Technologie.

De 2009 à 2013 : *Directeur du département informatique de la Faculté des Sciences et Techniques - Université de Corse, webmaster du site web dédié : <http://dptinfo.univ-corse.fr>.*

De 2011 à 2013 : *Correspondant ISN (Informatique et Sciences du Numérique) à l'Université de Corse pour le Rectorat de Corse : organisateur de la formation, planification, élaboration et animation de cours.*

En 2012 : *Participation à l'organisation du concours de technicien supérieur territorial pour le compte du centre de gestion de la fonction publique territoriale de Corse du Sud ; élaboration et correction de l'épreuve écrite, participation à l'épreuve orale et au jury final.*

De 2004 à 2012 : *Fonctions électives : conseils d'administration, de laboratoire et de faculté, commission des spécialistes et comité technique paritaire au sein de l'Université de Corse.*

- De 2008 à 2012 : Membre du conseil de la Faculté des Sciences et Techniques de l'université de Corse, élu dans le collège des Maîtres de Conférences.
- De 2009 à 2011 : Membre du comité technique paritaire de l'Université de Corse.
- De 2004 à 2007 : Membre de la commission de spécialistes "Mathématiques et Informatique", section 25 et 27, élu 2nd vice-président.
- De 2004 à 2007 : Membre du conseil de laboratoire SPE - UMR CNRS 6134, élu dans le collège des Maîtres de Conférences.
- De 2003 à 2007 : Membre du conseil d'administration de l'Université de Corse élu dans le collège des Maîtres de Conférences.

De 2002 à 2008 : *Vice doyen en charge de l'insertion professionnelle de la Faculté des Sciences et Techniques - Université de Corse.*

- De 2002 à 2008 : Créateur et coordinateur général du concours « 24 heures pour innover » visant à la sensibilisation des étudiants de master à l'entrepreneuriat et à l'innovation ; concours labélisé et référencé au niveau national par l'OPPE (Observatoire des Pratiques Pédagogiques en Entrepreneuriat) et l'APCE (Agence Pour la Création d'Entreprise), labélisé et sponsorisé au niveau régional (CTC, CCI, OSEO, DRRT, etc.) et parrainé chaque année par une entreprise insulaire du secteur privé.
- De 2004 à 2008 : Webmaster du portail internet de la Faculté des Sciences et Techniques de l'Université de Corse.
- De 2005 en 2006 : Dématérialisation de la gestion des stages étudiants en entreprise et des préinscriptions pour les filières sélectives de l'Université de Corse.

En 2007 : *Participation à l'organisation du concours de technicien supérieur territorial pour le compte du centre de gestion de la fonction publique territoriale de Haute-Corse ; élaboration et correction de l'épreuve écrite, participation à l'épreuve orale et au jury final.*

De 2005 à 2006 : *Responsable de la mention informatique de la licence sciences et technologies de la Faculté des Sciences et Techniques - Université de Corse.*

F. Synthèse des activités d'enseignements

(Pour plus de détails cf. section IV : Charges d'enseignement occupées)

Depuis 2013 : *Maître de Conférences en informatique au sein du département de Génie Informatique de Faculté d'Ingénierie et de Technologie de l'Université Galatasaray (GSÜ).*

- 6 heures de cours hebdomadaire : compilation, automates et théorie des langages, "complexity and information theory" et "information retrieval".
- Coordination de 7 enseignants français intervenant au sein des départements de génie informatique et industriel de la Faculté d'Ingénierie et de Technologie.

De 2002 à 2013 : *Maître de conférences en informatique à la Faculté des Sciences et Techniques de l'Université de Corse.*

- 192 heures par an équivalent TD : Gestion de projet informatique, algorithmique, structures de données et programmation, développement d'applications web, technologie d'échange de données : XML et web sémantique, intelligence artificielle et programmation concurrente.
- De 2011 à 2013 : Animation de cours dans le cadre de la formation « Informatique et Sciences du Numérique » (ISN) destinée à des enseignants du secondaire pour le compte du Rectorat de Corse.
- De 2011 à 2013 : Elaboration et animation d'un stage d'algorithmique pour les enseignants de Mathématiques du secondaire pour le compte du Rectorat de Corse ; base de l'algorithmique et initiation au langage python.
- En 2012 : Création et enseignement du cours « Algorithmique et Programmation » au sein de la spécialité énergétique de l'école d'ingénieur « Paoli Tech » : langage C, Matlab.
- De 2010 à 2012 : Membre et acteur du projet « Implementation of E-Learning Content for Energy Saving Farm into Vocational Education : ECEVE », No. 10310 0544, programme européen « LIFELONG LEARNING PROGRAMME » et de son action « LEONARDO DA VINCI Transfer of innovation », 280 k€, <http://eceve.org>.

- En 2010 : Enseignement du module B7 du C2i « Mener des projets en travail collaboratif à distance » dans le cadre de la formation continue du personnel de l'université de Corse.
- De 2008 à 2010 : Mise en place du tutorat relatif au « Plan Pécresse » pour la discipline informatique en 1^{ère} année de licence sciences, technologies et santé.
- En 2005 et 2006 : Concepteur et animateur du module B7 « Mener des projets en travail collaboratif à distance » du Certificat informatique et Internet (C2i) pour le compte du site web <http://w2.c2imes.org>.

G. Activités de consultance

Depuis 2009 : Auto-Entrepreneur : Activités d'expertises ou de consultations auprès de collectivités, d'entreprises et d'organismes privés ainsi que sur des prestations d'enseignements et de formations.

En 2012 : *Ellucian / SunGard Global Services : Groupe de 20.000 employés répondant aux besoins de plus de 25.000 clients, répartis dans plus de 70 pays à travers le monde.*

- Support métier aux experts fonctionnels et analyse de documents dans le cadre de la réponse à l'appel d'offre de l'université de Strasbourg.
- Formation des équipes sur le milieu universitaire français ayant permis aux consultants de mener à bien les ateliers fonctionnels liés au projet.

De 2007 à 2011 : *WMaker : Société de service spécialiste dans la conception de sites et d'applications web et mobiles.*

- Chef de projet pour l'outil collaboratif de veille et de publication « xFruits », <http://xfruits.com>.
- Participation au concours national d'aide à la création d'entreprises de technologies innovantes, Ministère délégué à l'Enseignement supérieur et à la Recherche avec le soutien de l'ANR et OSEO.
- Analyse et conduite des perspectives d'évolutions techniques et fonctionnelles.
- Suivi de la traduction de la documentation et du logiciel en plusieurs langues.
- Suivi des utilisateurs et des clients potentiels.
- Organisation de séminaires R&D : « web services », web sémantique.

H. Autres activités

Depuis 2000 : *Membre de l'IEEE, association professionnelle promouvant la connaissance dans le domaine de l'ingénierie électrique.*

- Depuis 2011 : Membre du CA et webmaster de la section France IEEE : <http://www.ieee-france.org>.
- En 2011 : Passage au grade de « IEEE Senior Member ».

Depuis 2005 : *Editeur francophone pour l'Open Directory Project Dmoz : plus grand annuaire web au monde.*

- Gestion de deux catégories avec près de 50 sites web ajoutés et maintenus : <http://www.dmoz.org>.

En 2006 et 2008 : *Développement de l'entrepreneuriat pour l'OFQJ / ACEE.*

- En 2008 : Sélectionné sur dossier pour une mission de prospection et de formation « Entrepreneuriat jeunesse » par l'OFQJ (Office Franco-Québécois pour la Jeunesse).
- En 2006 : Invité au 14^{ème} colloque annuel de l'ACEE (Association des clubs d'entrepreneurs étudiant) du Québec, « Entreprends au rythme du monde ! ».

En 2006 : *Partenariat entre Microsoft, le service pour la langue corse et la CTC.*

- Mise en place d'un partenariat entre Microsoft, la CTC (Collectivité Territoriale de Corse) pour proposer la traduction en langue corse de la suite bureautique, puis du système d'exploitation commercialisés par Microsoft.

En 2008 : *Stage linguistique « Comprehensive English » à l'Université de Toronto - Canada.*

- Stage intensif de 6 semaines.

En 2006 : *Livre blanc de la société de l'information en Corse.*

- Invité aux assises du livre blanc de la société de l'information en Corse pour présenter la politique de la Faculté des Sciences et Techniques de l'Université de Corse en matière d'insertion professionnelle (projet informatique, bourse de stage et emploi, dématérialisation des conventions de stage, etc.).

En 2006 : *Mise en place du site web PRODEnR.*

- Portail web d'information visant le renforcement des liens France/PECO (Pays d'Europe Centrale et Orientale) pour le développement des énergies renouvelables et soutenu par l'ADEME : <http://prodenr.univ-corse.fr>.

En 2006 : *Participation au projet NATURNET-REDIME.*

- Projet financé au titre du 6^{ème} programme cadre de recherche communautaire (FP6) : Exploitation du modèle de localisation participative d'un parc éolien en réalisant un scénario d'apprentissage pour le portail web de NATURNET-REDIME.

En 2006 : *Participation au pôle de compétitivité sur les EnR.*

- Participation à l'assemblée générale des acteurs de Corse Pôle Energies Renouvelables : pôle de compétitivité sur les EnR, Ajaccio.

De 2004 à 2006 : *Participation à la Fête de la science.*

- Présentation des thématiques de notre laboratoire en collaboration avec l'association "A Rinascita" (tête de réseau de la culture scientifique, technique et industrielle en Corse).

En 1998 : *Stage linguistique « English as a Second Language » l'Université du New Hampshire (UNH), Durham, NH - USA.*

- Stage intensif de 6 semaines, préparation au TOEFL.

III. Résumé des travaux de recherche et d'encadrement

La présente section propose de résumer mes activités en matière de recherche et d'encadrement en mettant en évidence leurs points forts, les activités de valorisation, les contrats de recherche, les développements de logiciels accomplis, les actions en matière d'information scientifique, technique, de vulgarisation et les activités internationales. Les aspects scientifiques seront développés plus précisément dans la section VII de ce rapport.

A. *Activités en matière de recherche*

Jusqu'à l'année 2004, mon travail de recherche s'est focalisé sur le test de systèmes électroniques. L'étude portait sur la génération automatique de vecteurs de test pour des circuits électroniques décrits à haut niveau d'abstraction (niveau algorithmique ou comportemental) avec le langage dédié VHDL (« VHSIC Hardware Description Language »). Suite à un changement d'équipe en 2005, j'ai eu l'opportunité de m'investir dans deux thématiques propres à l'étude des systèmes énergétiques à sources d'Energies Renouvelables (EnR) : la modélisation et la prédiction de séries temporelles environnementales en utilisant les techniques d'intelligence artificielle et la représentation des connaissances dans les processus d'aide à la décision. Ces trois thématiques se résument ainsi :

- **Validation et test de systèmes électroniques** : Notre but était d'aider le concepteur de circuits électroniques à tester ceux-ci au début de leur phase de conception à partir d'une description algorithmique. Du point de vue du test, une description de circuits électroniques au niveau comportemental (ou algorithmique) est similaire à un programme écrit avec un langage de programmation classique. Cela suggère que les techniques appliquées avec succès dans le domaine du test de logiciels peuvent se révéler efficaces pour tester les conceptions de type matériel écrites avec le langage VHDL. Un des principaux problèmes lié à ce type d'approche est la génération des vecteurs de test, qui sont formés des stimuli (valeurs à appliquer en entrée) et des réponses attendues sur les sorties d'un circuit électronique. Ce travail initié en 1998, essentiellement développé durant mes trois années de doctorat, a été poursuivi jusqu'en 2004.
- **Modélisation et prédiction de séries temporelles** : Nous nous sommes dans un premier temps intéressés à l'utilisation de réseaux de neurones artificiels (RNA) couplés à un prétraitement statistique des séries temporelles de rayonnement global. Puis nous avons étudié leur hybridation avec des méthodes de prédiction plus « classiques », tels que les modèles autorégressifs et moyenne mobile (ARMA). Ces techniques de prédiction trouvent une application au sein de la plateforme MYRTE (Mission hYdrogène Renouvelable pour l'inTégration au réseau Electrique) dans le but d'offrir une prévision à courte et très courte échéance de l'énergie produite afin d'optimiser la gestion du stockage hydrogène

dans les réservoirs et de piloter au plus fin le système énergétique en fonction de la charge à alimenter. Fort du succès rencontré pour la prédiction d'ensoleillement à différentes horizons (jour, heure, minute), l'application de ces techniques aux vitesses de vent et aux pics de pollution atmosphérique a également débutée. Concernant cette dernière, elle s'effectue dans le cadre d'un contrat de recherche avec Qualitair Corse, association agréée pour la surveillance de la qualité de l'air. Concernant la prédiction de vitesses de vent, un stage de Master a été effectué en 2013 en collaboration avec l'ENSAM de Paris et un sujet de thèse en co-tutelle avec ce même organisme a été déposé. Enfin nous avons démarré des travaux concernant l'estimation de potentiel solaire également à l'aide réseaux de neurones artificiels. Il s'agit cette fois non pas de prédire mais d'estimer le rayonnement solaire incliné à partir du rayonnement solaire global horizontal et d'autres variables météorologiques disponibles.

- **Représentation des connaissances dans les processus d'aide à la décision :** La recherche d'un processus d'aide à la décision efficient peut s'avérer particulièrement complexe et difficile lors du montage et la réalisation de projets de grande dimension, tels que les centrales photovoltaïques ou encore les fermes éoliennes. Trop souvent, la question du conflit et celle de l'acceptation sociale par la population de ce type de projets, est sous-estimée. Dans ce contexte nous pensons qu'un bon usage des outils et technologies informatiques, mais aussi que la construction de nouveaux logiciels doivent permettre de répondre à des impératifs de gouvernance plus transparente et plus participative particulièrement ancrée à l'échelon local. Nous proposons dans ce projet d'étudier l'apport des langages informatiques de représentation des connaissances afin de clarifier les processus d'aide à la décision au sein des méthodes d'Aide MultiCritère à la Décision (AMCD). Les méthodes d'AMCD, branche d'étude majeure de la recherche opérationnelle, sont des méthodes et des calculs permettant de choisir la meilleure solution ou la solution optimale parmi tout un ensemble de solutions. Leur utilisation plus forte dans le milieu non académique spécialiste est freinée par des difficultés de compréhension d'un vocabulaire très riche et complexe. Ces difficultés peuvent conduire à un manque d'intelligibilité des résultats pour les décideurs, voir à une défiance et une méfiance vis à vis de la méthode.

B. Points forts des activités de recherche

- **Validation et test de systèmes électroniques :** cette 1^{ère} thématique (mise en veille à partir de 2005), m'a permis de participer comme chercheur et responsable scientifique à deux contrats de recherche : (i) N°: F61775-00-C0002, projet 00-4005, item 0003 : « Validation of VHDL description », financé par l'EOARD : « European Office of Air Force Research and Development », ~15k € (2000-2003) ; (ii) « Développement d'un outil de test de circuit et amélioration de la testabilité des circuits », financé par la CTC « Collectivité Territoriale de Corse », ~16k € (2003-2004). L'approche originale de validation de description VHDL développée a permis la création d'un prototype logiciel appelé GENESI

(GENErator de Stimull) permettant de générer, à partir d'une description VHDL comportementale au niveau algorithmique, les vecteurs de test à appliquer sur une description VHDL plus poche du niveau matériel. Plusieurs publications dans des conférences internationales reconnues de rang A ont été acceptées.

- **Modélisation et prédiction de séries temporelles** : concernant cette 2^{ème} thématique, deux sujets de thèse ont pu être défini et initié, un contrat d'apprentissage de niveau master est en cours et de nombreux stagiaires français et étrangers de niveau master ont été encadrés. Ces études concernent l'utilisation de réseaux de neurones artificiels dans le cadre de la prédiction de données d'ensoleillement, de vitesses de vent mais aussi de pics de pollution atmosphérique.

Prédiction de données d'ensoleillement : un travail doctoral a débuté en janvier 2008 avec pour titre : « Prédiction de séries temporelles de rayonnement solaire global et de production d'énergie photovoltaïque à partir de réseaux de neurones artificiels ». L'objectif de cette thèse, soutenue en novembre 2011, a été de développer des techniques d'analyse et de reconstitution de séries temporelles tels que le rayonnement solaire sur un site afin d'en estimer le potentiel énergétique. Plusieurs publications dans des revues de rang A et publications dans des conférences internationales reconnues ont été acceptées. Ces travaux ont intéressé fortement l'opérateur EDF qui a mis à notre disposition en 2009 et 2010 deux stagiaires de niveau master sur cette thématique pour nous épauler lors de ce travail doctoral.

Nos travaux ont été référencés au sein de l'action COST (European Cooperation in Science and Technology) WIRE (Weather Intelligence for Renewable Energies) rassemblant différents acteurs du domaine dans le but d'améliorer les techniques de prévision météorologique dédiées aux énergies solaires et éoliennes.

Nous avons poursuivi ces travaux par la définition de modèles hybrides (réseaux de neurones artificiels et modèles linéaires de type ARMA) en les combinant grâce à des démarches dites d'inférences bayésiennes. Nous souhaitons plus largement établir une hiérarchie de différents prédictors selon l'horizon de prédiction considéré (jour, heure, minute) afin d'envisager la définition de stratégies innovantes de gestion intelligente de l'énergie.

La transposition de ces travaux de recherche dans le cadre de la plateforme MYRTE (Mission hYdrogène Renouvelable pour inTégration au réseau Electrique) et du smart-grid (PAGLIA ORBA) a été instruite dans le dépôt de dossier à l'ANR PROGELEC. Ce projet d'ANR cible la problématique de la prédiction de la production photovoltaïque, en collaboration avec nos collègues de l'Université de La Réunion, des Antilles-Guyane et de Polytechnique.

Prédiction de pics de pollution atmosphérique : Le projet d'étendre l'utilisation des réseaux de neurones artificiels à cette problématique a pu être initié en 2009. Un

contrat de recherche nous unit désormais à l'association agréée pour la surveillance de la qualité de l'air en Corse : « Qualitair Corse ». Deux stages de niveau master d'une période de 5 mois ont été accomplis en 2009 et 2010. Ces travaux ont fait l'objet d'une publication au sein d'une conférence internationale IEEE en 2011. Ils sont la base d'un projet plus ambitieux proposant de construire un prédicteur « h+24 » et pour lequel un travail doctoral a été défini. Nous souhaitons disposer un jour à l'avance, par exemple à 18h00 le jour J, les taux de pollutions du jour J+1 à chaque heure pour pouvoir prévenir les autorités en cas de prévision de dépassement des taux limites. Une demande de financement de doctorat de type CIFRE a été déposée conjointement auprès de la « Collectivité Territoriale de Corse » (CTC) sur fonds européens (CPER et FEDER) et de son Agence de Développement Economique de la Corse (ADEC). Cette demande a été acceptée et le doctorant a pu débuter ce travail en mai 2012 en tant que salarié au sein de Qualitair Corse.

Les acteurs de ces travaux sont impliqués au sein de l'observatoire CORSICA (« Centre d'Observation Régional pour la Surveillance du Climat et de l'environnement Atmosphérique et océanographique en Méditerranée occidentale ») et du chantier MISTRALS (« Mediterranean Integrated STudies at Regional And Local Scales »), de l'Alliance ALLENI et de son axe ChArMEx (« the Chemistry-Aerosol Mediterranean Experiment »), initiative de l'INSU.

Il est à noter que notre université et l'IESC de Cargèse ont eu l'honneur d'organiser le 3^{ème} workshop international ChArMEx en novembre 2012 et de recevoir l'ensemble des membres de cette communauté, en intégrant le comité scientifique, en présidant le comité d'organisation et en animant le site web de cette manifestation : <http://charmex.lsce.ipsl.fr>.

- **Représentation des connaissances dans les processus d'aide à la décision** : Concernant cette 3^{ème} thématique, un sujet de thèse a été défini et initié et un prototype logiciel est en fin de développement. Un projet nommé MECADEPPE (MEdiation des connaissances d'acteurs destinée à l'Evaluation des projets et politiques d'Energie) est en cours d'instruction au sein du pôle de compétitivité « Energies non génératrices de gaz à effet de serre ». Le projet a reçu un avis favorable des deux conseils scientifiques (Corse et PACA). Un article dans un ouvrage collectif de référence et plusieurs communications ont été acceptés. De plus, nous avons intégré le projet européen « Decision-Deck » dont l'objectif est de développer une plateforme logicielle générique d'aide multicritère à la décision. Nous y étudions l'apport de langages informatiques dédiés pour représenter la connaissance produite tout au long du processus de décision. Il est à noter que notre université a eu l'honneur d'organiser les 73^{èmes} journées de l'AMCD (Aide MultiCritère à la Décision) couplées aux 8^{èmes} journées de travail du projet « Decision Deck » en avril 2011 et de recevoir l'ensemble des membres de cette communauté, en intégrant le comité scientifique, en présidant le comité d'organisation et en animant le site web de ces deux manifestations : <http://mcda.univ-corse.fr>.

Concernant le travail doctoral, il s'agit d'un co-encadrement sur le sujet suivant :
« Evaluation multicritère et e-gouvernance territoriale des ressources naturelles au Cameroun » avec l'Université de Ngaoundéré (Cameroun), Département de Mathématiques et Informatique, démarrée en juin 2012.

C. *Valorisation : contrats, dépôts de brevets, logiciels*

- Depuis 2013 : Participation au développement d'un logiciel de visualisation et de traitement de données multi-sources liée à la prédiction de pics de pollution pour le compte de l'association agréée Qualitair Corse.
- Depuis 2008 : Membre du projet européen « Decision-Deck » dont l'objectif est de développer une plateforme logicielle générique d'aide multicritère à la décision.
- Depuis 2008 : Développement d'un prototype logiciel monoposte mono-utilisateur dans le cadre du projet MECADEPPE (MEdiation des Connaissances d'Acteurs Destinée à L'Evaluation des Projets et Politiques d'Energie).
- De 2005 à 2011 : Participation à l'évolution du système de gestion de contenu (CMS en anglais pour "Content Management System") développé par la société WMaker : Conseil, test, proposition de séminaires, veille technologique, dépôt d'un sujet de thèse en co-tutelle avec l'INRIA de Nice Sophia Antipolis.
- En 2008 : Dépôt du projet MECADEPPE auprès du pôle de compétitivité « Energies non génératrices de gaz à effet de serre ».
- En 2007 : Participation au concours national d'aide à la création d'entreprises de technologies innovantes, Ministère délégué à l'Enseignement supérieur et à la Recherche avec le soutien de l'ANR et OSEO. Xfruits, <http://www.xfruits.com> : Outil collaboratif de veille et de publication.
- En 2006 : Mise en place d'un partenariat entre Microsoft, le service pour la langue corse et la Collectivité Territoriale de Corse (CTC) pour proposer la traduction en langue corse de la suite bureautique, puis du système d'exploitation commercialisés par Microsoft.
- De 2001 à 2004 : Développement du prototype logiciel GENESI (GENERateur de Stimuli) permettant de générer des stimuli à partir d'une description VHDL comportementale au niveau algorithmique.

D. *Information scientifique, technique et vulgarisation*

- En 2013 : Participation au numéro spécial « Energie » de « Stantari », revue trimestrielle de vulgarisation scientifique, traitant du patrimoine naturel et historique corse : « Prédire le vent et le soleil en s'inspirant du fonctionnement du cerveau ».
- Depuis 2011 : Participation à l'organisation et l'animation de la formation « Informatique et Sciences du Numérique » (ISN) destinée à des enseignants

du secondaire pour le compte du Rectorat de Corse : élaboration et animation de cours, participation au jury.

- En 2010 : Rapport interne, « Le C2i à l'Université de Corse - Pasquale PAOLI : Harmonisation, organisation du niveau 1 et projet de niveau 2 », T. Antoine-Santoni, C. Paoli.
- En 2008 : Organisation d'un séminaire pour la société Webzine Maker : Le web sémantique, URI, XML, RDF, OWL, SPARQL, et ses implications économiques, avril 2008, Ajaccio, France.
- De 2004 à 2006 : Participation à la Fête de la science, présentation des thématiques de notre laboratoire.
- En 2006 : Rapport interne, « Application web pour la mise en œuvre de démarches d'AMCDP », C. Paoli, M-L. Nivet, P. Oberti, 2006, Corte, France.
- En 2006 : Mise en place du site web PRODEnR pour renforcer les liens France/PECO (Pays d'Europe Centrale et Orientale) soutenu par l'ADEME pour le développement des énergies renouvelables : <http://prodenr.univ-corse.fr>.
- En 2005 : Organisation d'un séminaire pour la société Webzine Maker : Présentation de la technologie des Web Services, présentation générale et technique, effectuée en collaboration avec M-L. NIVET, août 2005, Corte, France.
- En 2005 : Participation au projet NATURNET-REDIME, financé au titre du 6^{ème} programme cadre de recherche communautaire (FP6) : Exploitation du modèle de localisation participative d'un parc éolien en réalisant un scénario d'apprentissage pour le portail NATURNET-REDIME.

E. Activités internationales : conférences invitées, contrats, séjours à l'étranger, accueil de collègues et stagiaires étrangers

- Depuis 2013 : Nommé Expert Technique International (ETI) par le Ministère des Affaires Etrangères et du Développement International à l'université Galatasaray (Turquie) en tant qu'enseignant chercheur en informatique au sein du département de Génie Informatique de Faculté d'Ingénierie et de Technologie.
- Depuis 2013 : Participation au programme Hubert Curien (PHC) franco-grec PLATON, projet 30355UE, développement de modèles durables d'approvisionnement en énergie et en eau pour les îles Méditerranéennes par utilisation des EnR.
- De 2010 à 2012 : Membre et acteur du projet « Implementation of E-Learning Content for Energy Saving Farm into Vocational Education : ECEVE », No. 10310 0544, programme européen « LIFELONG LEARNING PROGRAMME » et de son action « LEONARDO DA VINCI Transfer of innovation ». D'une durée de 24 mois, ce projet a pour principal objectif la création de contenus interactifs dans le domaine des ressources énergétiques renouvelables, des procédures visant à réduire l'intensité énergétique des fermes agricoles et du changement climatique dû aux émissions de gaz à effet de serre.

- Depuis 2010 : Membre du comité scientifique et référés pour la conférence internationale IEEE Conference on Environment and Electrical Engineering (EEEIC) : <http://eeeic.eu>.
- Depuis 2006 : Encadrement et co-encadrement de plusieurs stages de master du réseau PECO (Pays d'Europe Centrale et Orientale) soutenu par l'ADEME (pour plus de détails cf. section v : Fonctions d'intérêt collectif exercées).
- En 2011 : Accueil de M. Sikorski de Warsaw University College of Technology and Business (WUCTB) dans le cadre d'une mobilité d'enseignement Erasmus pour des cours d'informatique en master.
- En 2008 : Participation à la conférence « Open Scholarship: Authority, Community and Sustainability in the Age of Web 2.0 », Knowledge Media Design Institute (KMDI) de l'Université de Toronto, Canada.
- En 2008 : Stage intensif linguistique « Comprehensive English » de 6 semaines à l'Université de Toronto, Canada.
- En 2008 : Dossier retenu par l'OFQJ (Office Franco-Québécois pour la Jeunesse) pour participer à une Mission de prospection et de formation « Entrepreneuriat jeunesse » à Montréal et Québec, Canada, 10 au 19 juin 2008.
- En 2007 : Visite de la Faculté de Sliven, Bulgarie dans le cadre de l'organisation européenne Socrates/Erasmus.
- En 2006 : Invité au 14^{ème} colloque annuel de l'ACEE du Québec, « Entreprends au rythme du monde! », Rivière du Loup, Canada.
- En 1998 : Stage linguistique « English as a Second Language » l'Université du New Hampshire (UNH), Durham, NH - USA, Stage intensif de 6 semaines, Préparation au TOEFL.

IV. Charges d'enseignement occupées

Mes activités d'enseignement se déroulent aujourd'hui à l'Université Galatasaray en Turquie où j'ai été nommé en 2013 Expert Technique International (ETI) par le Ministère français des Affaires Etrangères et du Développement International. J'y exerce au sein du département de Génie Informatique de la Faculté d'Ingénierie et de Technologie les enseignements suivants : compilation, automates et théorie des langages, complexité et théorie de l'information ("Complexity and Information Theory") et recherche d'information ("Information Retrieval").

Avant cette période, la majeure partie de mes activités d'enseignement s'est déroulée à la Faculté des Sciences et Techniques (FST) de l'Université de Corse et plus précisément au sein du département informatique dont j'ai été le directeur de 2009 à 2013. Deux diplômes sont intégralement gérés au sein de ce département : une 3^{ème} année de licence généraliste en informatique et un master mention informatique spécialité « Systèmes d'Information et Internet ». Mes enseignements couvraient les domaines suivants : gestion de projet informatique, algorithmique, structures de données et programmation, développement d'applications web, technologies d'échange de données (XML et web sémantique), intelligence artificielle et programmation concurrente. Outre ces enseignements gérés dans ce département, j'ai également dispensé des cours en 1^{ère} et 2^{ème} année de licence généraliste (algorithmique et programmation) ; en 1^{ère} année de la spécialité énergétique de l'école d'ingénieur « Paoli Tech » (algorithmique et programmation) ; dans le cadre de la formation continue du personnel de l'Université de Corse (Certificat Informatique et Internet) ; et pour le compte du Rectorat de Corse (Informatique et Sciences du Numérique).

De plus, j'ai par deux fois participé à l'organisation du concours de technicien supérieur territorial pour le compte du centre de gestion de la fonction publique territoriale de Haute Corse puis de Corse du Sud : élaboration et correction de l'épreuve écrite, participation à l'épreuve orale et au jury final.

Enfin, j'ai participé au Programme Leonardo da Vinci – Transfert de Technologie, pour lequel j'ai produit du contenu de formation numérique. Le projet intitulé ECEVE (E-Learning Content for Energy Saving Farm in Vocational Education) a été réalisé en coopération avec les établissements suivants : Slovak University of Agriculture, Agroinstitut (Slovaquie), Multimedia Sunshine Ltd (UK), Infoart (Bulgarie), Kahramanmaraş Technical and Vocational High School (Turquie), Czech Technical University (République Tchèque).

A. *Université Galatasaray (GSÜ)*

Fondée en 1992, l'Université Galatasaray (GSÜ) est une université publique turque francophone établie à Istanbul (Turquie). Mon contrat prévoit un service hebdomadaire de 6 heures de cours réparti sur 2 semestres de 14 semaines.

Les formations supérieures turques reposent sur un système de notation et de crédits calqué sur le système américain. Avec le développement des échanges Erasmus, GSÜ a adopté le système de transfert de crédits européens (ECTS). Par contre, il n'y a pas pour l'instant de travaux sur le rapprochement avec le système LMD. Ainsi en Turquie les études supérieures s'organisent en trois cycles principaux : la « Lisans » en 4 ans (équivalente à l'ancienne Maîtrise française), le Master ou « Lisans » supérieure en 2 ans (équivalente à notre Master ou aux anciens DEA et DESS français), le Doctorat (correspondant au Doctorat français).

Les enseignements que j'anime peuvent être listés comme suit :

Lisans généraliste d'informatique : 3^{ème} et 4^{ème} année

- Depuis 2013 : Compilation (3H par semaine) au 1^{er} semestre
- Depuis 2013 : Automates et théorie des langages (3H par semaine) au 2nd semestre

Master d'informatique : 1^{ère} année

- Depuis 2013 : "Complexity and Information Theory" (3H par semaine) au 1^{er} semestre
- Depuis 2013 : "Information Retrieval" (3H par semaine) au 2nd semestre

A noter que pour renforcer l'attractivité du Master d'informatique, le département de Génie Informatique a décidé que les cours de ce diplôme seraient dispensés en langue anglaise, les cours de Lisans restant dispensés en langue française.

B. *Université de Corse - Faculté des Sciences et Techniques (FST)*

Ma carrière d'enseignant a débuté en 1998 en qualité de doctorant-moniteur (jusqu'en 2001). Statutairement seuls des TD et des TP de 1^{er} cycle universitaire pouvaient m'être attribués. J'ai ainsi pu initier des étudiants de ce cycle aux logiciels de bureautique, à l'algorithmique et à la programmation procédurale.

A partir de 2001 en qualité d'ATER puis de Maître de Conférences, j'ai pu dispenser des cours en 1^{er} et 2^{ème} cycle. J'ai eu ainsi la charge d'organiser plusieurs Unités d'Enseignement (UE). Chacun de ces enseignements a donné lieu à la création d'un support de cours, des planches de TD et de TP numérisés et distribués en début de semestre et accessibles sur le web : sur mes pages web personnelles, puis sur l'environnement numérique de travail (ENT) de l'Université de Corse : <http://campus.univ-corse.fr>.

L'évaluation des connaissances et des compétences est réalisée soit par un contrôle continu et régulier, soit par un examen terminal, soit par ces deux modes. Les annales des examens précédents sont également mises à disposition.

Enfin, je me suis fortement impliqué concernant le suivi des stages et plus particulièrement en 2nde année de master : aide à la recherche de stage (CV, lettre de motivation, réseautage), aide à la rédaction du rapport et à la préparation de l'oral, participation au jury, tutorat de stage. On peut évaluer à une moyenne de 2, le nombre de stagiaires tuteurés par an.

Les enseignements que j'ai pu animer ou auxquels j'ai participé peuvent être listés comme suit :

Licence généraliste : 1^{ère} et 2^{ème} année

- 2001 à 2008 : Fondamentaux de l'informatique (18 CM / 18 TD / 18 TP).
- 2001 à 2007 : Introduction à l'algorithmique et à la programmation (18 CM / 18 TD / 18 TP).
- 2004 à 2009 : Algorithmique et structure de données I et II (2 * {18 CM / 18 TD / 18 TP}).
- 2008 à 2010 : De la puce au web (18 CM / 18 TD / 18 TP) avec mise en place du tutorat relatif au « Plan Pécrasse ».

Licence généraliste d'informatique : 3^{ème} année

- 2004 à 2006 : Programmation Java (36H eq. TD).
- 2005 à 2013 : Algorithme des graphes (18 CM / 18 TD / 18 TP).
- 2007 à 2013 : Technologies d'accès aux données, technologies XML et programmation web (18 CM / 18 TD / 18 TP).
- 2010 à 2013 : Concepts des langages de programmation, programmation concurrente (18H eq. TD).

Master d'informatique : 1^{ère} et 2nde année

- 2005 à 2013 : Gestion de projets informatiques (36H eq. TD).
- 2007 à 2013 : Représentation de l'info pour le Web : le web sémantique (18 CM / 18 TD).
- 2009 à 2011 : Intelligence artificielle : les réseaux de neurones artificiels (18H eq. TD).

C. Université de Corse - Ecole d'Ingénieur « Paoli Tech »

Ouverte en 2011, la spécialité « Energétique » de l'école d'ingénieur « Paoli Tech » propose à ses élèves ingénieurs des enseignements visant la maîtrise de l'outil informatique, de la programmation et de l'algorithmique. Dans ce contexte j'ai créé et enseigné le cours suivant :

1^{ère} année de la spécialité énergétique

- 2011 : Algorithmique et programmation : langage C, Matlab (9 CM / 9 TD / 9 TP).

D. Université de Corse - Formation Continue

Soucieuse du bon usage par son personnel administratif et enseignant de l'outil informatique, l'Université de Corse a mis en place une formation au Certificat Informatique et Internet (C2i). J'ai ainsi été sollicité par deux fois pour animer le module B7 du C2i dont j'ai conçu un support numérique :

Formation au C2i du personnel de l'Université de Corse

- En 2010 et 2011 : Enseignement du module B7 du C2i « Mener des projets en travail collaboratif à distance » enseigné dans le cadre de la formation continue du personnel de l'université de Corse.
- En 2005 et 2006 : Concepteur et animateur du module B7 « Mener des projets en travail collaboratif à distance » du C2i pour le compte du portail web : <http://w2.c2imes.org>.

E. Rectorat de Corse

Le Rectorat de Corse a sollicité le département informatique de la FST pour accompagner la mise en place du programme d'Informatique et Sciences du Numérique (ISN) en classe terminale de la série scientifique au lycée. J'ai ainsi pu coordonner en totalité, et animer en partie cette action de formation.

Organisation de la formation « Informatique et Sciences du Numérique » (ISN)

- En 2012 : Organisation de la formation « Informatique et Sciences du Numérique » (ISN) « avancée » destinée aux enseignants du secondaire habilités pour le compte du Rectorat de Corse ; planification, élaboration et animation de cours.
- En 2011 : Organisation de la formation « Informatique et Sciences du Numérique » (ISN) destinée à des enseignants du secondaire pour le compte du Rectorat de Corse ; planification, élaboration et animation de cours, participation au jury.

- En 2011 : Stage d'algorithmique pour les enseignants de Mathématiques du secondaire pour le compte du Rectorat de Corse ; base de l'algorithmique et initiation au langage python.

F. Centre départemental de gestion de la fonction publique territoriale

Le département informatique de la FST a été sollicité par deux fois pour participer à l'organisation du concours de technicien supérieur territorial. J'ai ainsi pu collaborer à l'élaboration et à la correction de l'épreuve écrite, participer à l'épreuve orale et au jury final de ces épreuves.

Participation à l'organisation du concours de technicien supérieur territorial

- En 2012 : Elaboration et correction de l'épreuve écrite, participation à l'épreuve orale et au jury final, pour le compte du centre de gestion de la fonction publique territoriale de Corse du Sud.
- En 2007 : Elaboration et correction de l'épreuve écrite, participation à l'épreuve orale et au jury final, pour le compte du centre de gestion de la fonction publique territoriale de Haute-Corse.

V. Fonctions d'intérêt collectif exercées

Les fonctions d'intérêts collectif que j'ai pu exercer sont listés en deux parties : responsabilités scientifiques (encadrement de thèses, de masters, projets de recherche, référés, etc.) et responsabilités administratives (commission de spécialistes, charges électorales, organisation de conférences, responsabilité de diplômes, direction de faculté, etc.).

A. Responsabilités scientifiques

Développés au sein du projet « Energie Renouvelable » du laboratoire Sciences Pour l'Environnement (SPE) - UMR CNRS 6134, mes travaux de recherche se concentrent sur deux domaines : (i) l'utilisation des techniques d'intelligence artificielle pour la modélisation et la prédiction de séries temporelles et (ii) l'étude de langages informatiques de représentation des connaissances dans les processus d'aide à la décision. Je possède également une expérience, acquise au cours de mon doctorat, dans le domaine du test de logiciel appliqué à des circuits électroniques décrits en langage VHDL « VHSIC Hardware Description Language ».

1. Encadrements de thèses

- Depuis 2012 : Co-direction de thèse de M. Frédéric GUIDANA GAZAWA : « Evaluation multicritère et e-gouvernance territoriale des ressources naturelles au Cameroun », direction du Professeur Paul Henri AMVAM ZOLLO, département de Mathématiques et Informatique, Université de Ngaoundéré (Cameroun),.
- Depuis 2012 : Co-direction de thèse de M. Wani TAMAS : « Prévion des concentrations de polluants atmosphériques en Corse et impacts sur la productivité des systèmes énergétiques solaires », direction de M. Gilles NOTTON, Université de Corse, bourse de type CIFRE.
- De 2008 à 2011 : Co-direction et membre du jury de thèse de M. Cyril VOYANT : « Prédiction de séries temporelles de rayonnement solaire global et de production d'énergie photovoltaïque à partir de réseaux de neurones artificiels », direction de M. Marc MUSELLI, Université de Corse, qualifié CNU 62.

2. Participation à des encadrements de thèses

- Depuis 2013 : Participation aux travaux de Mlle Kahina DAHMANI : « Utilisation des réseaux de neurones artificiels pour l'estimation du potentiel solaire en Algérie », co-direction de M. G. NOTTON, Université de Corse et de M. Rabah DIZENE, Laboratoire de Mécanique Avancée, Université des

Sciences et de la Technologie Houari Boumediene (USTHB), Algérie, thèse démarrée en 2012.

- De 2010 à 2011 : Participation aux travaux de M. Ionut CALUIANU « Augmentation de la productivité énergétique des modules photovoltaïques », co-direction de M. G. NOTTON, Université de Corse et de Mme I. COLDA, Université Technique de Construction de Bucarest, thèse soutenue le 06 septembre 2011 à l'Université Technique de Construction de Bucarest, Roumanie, mention exceptionnelle.

3. Encadrements de master

- Depuis 2012 : Direction scientifique du contrat d'apprentissage d'une durée de 2 ans de Mlle Aurelia BALU en master informatique : « Techniques de l'intelligence artificielle et développement logiciel dans le cadre de la prédiction de séries temporelles ».
- En 2013 : Co-encadrement de M. W ZHAO pour un stage de 5 mois, « Prediction of wind energy capacity using artificial intelligence methods », école des Arts et Métiers ParisTech.
- En 2012 : Co-encadrement de M. N.A. KLIMOV pour un stage de 5 mois, « Utilisation des réseaux neuronaux pour l'estimation des données d'irradiation solaire directe », stage de fin d'étude, Université Technique de Sofia.
- En 2012 : Encadrement de M. M. MAZOUZ pour un stage de 5 mois « Développement d'un plugin logiciel et intégration au sein d'une plate-forme générique d'aide à la décision », Université de Corse.
- En 2011 : Co-encadrement de Mlle L. IVANOVA pour un stage de 5 mois, « Utilisation des réseaux neuronaux pour l'estimation des données d'irradiation solaire sur des plans inclinés », stage de fin d'étude, Université Technique de Sofia.
- En 2011 : Co-encadrement de Mlle S. VASILEV pour un stage de 5 mois, « Application des réseaux neuronaux à l'estimation du rayonnement diffus » stage de fin d'étude, Université Technique de Sofia.
- En 2010 : Co-encadrement de M. M. PADOVANI pour un stage de 5 mois, « Mise en place d'un prédicteur de pics de pollutions atmosphériques à partir de réseaux neuronaux », stage de fin de deuxième année, ENSAM ParisTech, Aix en Provence.
- En 2009 : Encadrement de M. JB. MARCHETTI pour un stage de 5 mois, « Mise en œuvre d'une application informatique visant le développement et l'aide à la localisation multicritère de projets photovoltaïques en Corse », Université de Corse.
- En 2007 : Encadrement de Mlle S. GANCHEVA pour un stage de 5 mois, « Comparaison entre différentes méthodes pour la prédiction de vitesses de vent : Réseaux de Neurones Artificiels et modèles ARMA », stage de fin d'étude, Université technique de Sofia.

- En 2006 : Encadrement de Mlle V. ANDRONOVA pour un stage de 5 mois, « Utilisation de données météo et des réseaux de neurones pour la prédiction de vitesses de vent », stage de fin d'étude, Université technique de Sofia.

4. Participation à des encadrements de master

- En 2010 : Participation aux travaux de Mlle M. BATTISTINI pour un stage de 5 mois, « Etude statistique pour l'élaboration d'un modèle de prévision de l'indice de la qualité de l'air en Corse », Ecole des Mines de Douai.
- En 2010 : Participation aux travaux de Mlle P. RANDIMBIVOLOLONA pour un stage de 6 mois, « Prédiction du rayonnement solaire global horizontal à l'horizon H+24 », Mines ParisTech.
- En 2009 : Participation aux travaux de M. M. UGER pour un stage de 6 mois, « Intérêt des techniques de stockage pour le développement des ENR à caractère aléatoire », Mines ParisTech.

5. Participation à des projets scientifiques

- Depuis 2013 : Participation au programme Hubert Curien (PHC) franco-grec PLATON, projet 30355UE, développement de modèles durables d'approvisionnement en énergie et en eau pour les îles Méditerranéennes par utilisation des EnR.
- Depuis 2009 : Chercheur au sein de la plateforme MYRTE (Mission hYdrogène Renouvelable pour l'inTégration au réseau Electrique), ~21 M € : prévision à courte et très courte échéance de l'énergie produite dans le but d'optimiser la gestion du stockage hydrogène dans les réservoirs et de piloter au plus fin le système énergétique en fonction de la charge à alimenter ; co-animateur du site web dédié : <http://myrte.univ-corse.fr>.
- Depuis 2009 : Chercheur et responsable scientifique au sein de l'observatoire CORSICA (« Centre d'Observation Régional pour la Surveillance du Climat et de l'environnement Atmosphérique et océanographique en Méditerranée occidentale ») ~2,2 M € : prévision des concentrations de polluants atmosphériques en Corse.
- Depuis 2010 : Dépositaire et référent d'un contrat de collaboration de recherche avec l'association agréée Qualitair Corse. Les objectifs de ce projet sont d'effectuer une étude complète sur le thème de la prévision de la qualité de l'air, et d'étudier la faisabilité d'un outil statistique et informatique de prévision.
- Depuis 2009 : Coordinateur local pour l'UMR CNRS 6134 du chantier MISTRALS (« Mediterranean Integrated STudies at Regional And Local Scales »), de l'Alliance ALLENI et de son axe ChArME (« the Chemistry-Aerosol Mediterranean Experiment »), initiative de l'INSU.

- Depuis 2008 : Intégration et participation au projet européen « Decision-Deck » dont l'objectif est de développer une plateforme logicielle générique d'aide multicritère à la décision, membre du consortium éponyme, <http://www.decision-deck.org>.
- De 2010 à 2012 : Membre et acteur du projet « Implementation of E-Learning Content for Energy Saving Farm into Vocational Education : ECEVE », No. 10310 0544, programme européen « LIFELONG LEARNING PROGRAMME » et de son action « LEONARDO DA VINCI Transfer of innovation », 280 k€, <http://eceve.org>.
- De 2003 à 2004 : Responsable scientifique et coordinateur du contrat de recherche « Développement d'un outil de test de circuit et amélioration de la testabilité des circuits », financé par la CTC « Collectivité Territoriale de Corse », ~16k €.
- De 2000 à 2003 : Participation comme chercheur au contrat de recherche international N°: F61775-00-C0002, projet 00-4005, item 0003 : « Validation of VHDL description », financé par l'EOARD : « European Office of Air Force Research and Development », ~15k €.
- En 1999 : Programme d'échange international (stage de 3 mois) au sein du laboratoire « Electrical and Computer Engineering » de l'UNH, « University of New Hampshire » (Durham, NH, USA) sur le projet : « Overview of Test Methods Using Boundary Scan ».

6. Référés et conférences internationales

- Depuis 2012 : Référés pour la revue « Renewable Energy » (Elsevier) : <http://ees.elsevier.com/rene/> (IF2011 2,978; IF5 ans 3,200, Hindex:60).
- Depuis 2010 : Membre du comité scientifique et référés pour la conférence IEEE « International Conference on Environment and Electrical Engineering » : <http://eeeic.eu>.
- En 2012 : Président du comité d'organisation et webmaster du 3^{ème} workshop international ChArMEx, 5-8 novembre 2012 à l'IESC de Cargèse, (webmaster du site web dédié : <http://charmex.lsce.ipsl.fr>).
- En 2011 : Président du comité d'organisation et membre du comité scientifique de la 8^{ème} édition du « Decision Deck Workshop » et de la 73^{ème} réunion du groupe de travail européen « Aide Multicritère à la Décision », (webmaster du site web dédié : <http://mcda.univ-corse.fr>).
- De 1998 à 2000 : Participation aux différentes éditions du workshop international ATW (« Advance Technology Workshop ») organisé par l'université de Corse et sponsorisé par la CTC et l'US Air force : conférence pluridisciplinaire proposant de mettre en synergie universitaires et industriels.

7. Fonctions électives

- De 2004 à 2007 : Membre de la commission de spécialistes N°5 - Mathématiques et Informatique section 25 et 27, élu 2nd vice-président.
- De 2004 à 2007 : Membre du conseil de laboratoire - SPE UMR CNRS 6134 - élu dans le Collège des Maîtres de Conférences.

8. Divers

- Depuis 2010 : Référencement de nos travaux en matière de prédiction au sein de l'action COST (« European Cooperation in Science and Technology ») WIRE (« Weather Intelligence for Renewable Energies ») rassemblant différents acteurs du domaine dans le but d'améliorer les techniques de prévision météorologique dédiées aux énergies solaires et éoliennes.
- Depuis 2010 : Passage au grade de « IEEE Senior Member », membre du Conseil d'Administration, webmaster de la section France : <http://www.ieeefrance.org>.

B. Responsabilités administratives

Avec plus de 10 ans d'expérience au sein de l'Université de Corse, mon champ d'expertise inclut également une forte compétence administrative acquise comme vice-doyen chargé de l'insertion professionnelle (création du concours de sensibilisation à l'entrepreneuriat et à l'innovation « 24h pour innover », relation avec les entreprises, collecte de la taxe d'apprentissage, etc.) ; puis comme responsable de diplôme et directeur du département informatique (politique de formation et de recrutement, gestion des heures et répartition des services, gestion des salles de travaux pratiques, etc.). Enfin j'ai plusieurs fois été élu et siégé dans différents conseils et comités de l'Université de Corse.

1. Directions

- **De 2002 à 2008 : Vice-doyen en charge de l'insertion professionnelle.**
 - Créateur et coordinateur général du concours « 24 heures pour innover » visant à la sensibilisation des étudiants de master à l'entrepreneuriat et à l'innovation ; concours labélisé et référencé au niveau national par l'OPPE (Observatoire des Pratiques Pédagogiques en Entrepreneuriat) et l'APCE (Agence Pour la Création d'Entreprise), labélisé et sponsorisé au niveau régional (CTC, CCI, OSEO, DRRT, etc.) et parrainé chaque année par une entreprise insulaire du secteur privé.
 - Organisation de rencontres avec les professionnels : forum des métiers de l'informatique, forum des métiers du sport.

- Etablissement de partenariats avec les entreprises et institutions : comité de parrainage, conseil de perfectionnement.
 - Collecte de la taxe d'apprentissage : organisation de la campagne de communication, édition et création de documents disponibles sur internet, liens avec les organismes collecteurs, etc.
 - Edition de statistiques, réseautage social et interviews permettant le suivi des anciens étudiants.
- **De 2004 à 2008 : Webmaster du portail internet de la Faculté des Sciences et Techniques.**
 - 10.000 visites uniques mensuelles.
 - Diffusion d'une newsletter bimensuelle à près de 4000 abonnés.
 - Animation d'un blog professionnel informant des tendances du marché et des derniers articles parus dans la presse spécialisée.
 - Animation d'une bourse aux stages et aux emplois : 150 offres publiées par an.
 - Mise en ligne d'outils d'aides à la recherche de stages, emplois et jobs d'été.
 - **De 2005 à 2006 : Dématérialisation de la gestion des stages étudiants en entreprise et des préinscriptions pour les filières sélectives.**
 - Création d'applications web dédiées et généralisées à l'ensemble de l'université.
 - Prix spécial de l'innovation territoriale en 2006 pour l'application de gestion des stages.
 - Près de 800 conventions traitées, édition de statistiques concernant nos partenaires.
 - Mise à disposition de données concernant l'insertion professionnelle de nos étudiants lors du débat national université emploi.
 - Plus de 600 préinscriptions traitées par an.
 - Mise à disposition de données et édition de statistiques pour la scolarité centrale et les responsables de diplômes.

2. Fonctions pédagogiques

- **De 2009 à 2013 : Directeur du département informatique.**
 - Politique de recrutement et de formation.
 - Gestion des heures et répartition des services.
 - Pédagogie, planification des cours, gestion des salles de travaux pratiques.
 - Politique en matière de stage, projet professionnel et alternance.
 - Réseautage social, suivi et services aux étudiants.
 - Webmaster du site dédié : <http://dptinfo.univ-corse.fr>.

- **De 2005 à 2006 : Responsable de la mention informatique de la licence sciences et technologies.**
 - Gestion des emplois du temps.
 - Campagne d'habilitation.
 - Sélection des étudiants.

3. Fonctions électives

- De 2008 à 2012 : Membre du conseil de la Faculté des Sciences et Techniques (FST) de l'université de Corse, élu dans le collège des maîtres de conférences.
- De 2009 à 2011 : Membre du Comité Technique Paritaire de l'Université de Corse.
- De 2003 à 2007 : Membre du Conseil d'Administration de l'Université de Corse élu dans le collège des maîtres de conférences.

VI. Liste des publications

A. Publications dans des revues internationales à comité de lecture de Rang A

- K. Dahmani, R. Dizene, G. Notton, **C. Paoli**, C. Voyant, M-L. Nivet, « Estimation of 5-min time-step data of tilted solar global irradiation using ANN (Artificial Neural Network) model », *Energy*, Volume 70, Pages 374-381, June 2014 (IF2011 3,487; IF5 ans 3,858, HIndex:60).
<http://dx.doi.org/10.1016/j.energy.2014.04.011>
- C. Voyant, P. Haurant, M. Muselli, **C. Paoli**, M-L. Nivet, « Time series modeling and large scale global solar radiation forecasting from geostationary satellites data » *Solar Energy*, Volume 102, Pages 131-142, 2014 (IF2012 2.952; IF5 ans 3,433, HIndex:72).
<http://dx.doi.org/10.1016/j.solener.2014.01.017>
- C. Voyant, C. Darras, M. Muselli, **C. Paoli**, M-L. Nivet, P. Poggi, « Bayesian rules and stochastic models for high accuracy prediction of solar radiation », *Applied Energy* 114, 218-226, 2014 (IF2012 4,781; IF5 ans 4,783, HIndex:53).
<http://dx.doi.org/10.1016/j.apenergy.2013.09.051>
- C. Voyant, **C. Paoli**, M. Muselli, M-L. Nivet, « Multi-horizon solar radiation forecasting for Mediterranean locations using time series models », *Renewable and Sustainable Energy Reviews* 28, 44-52, 2013 (IF2012 5,627; IF5 ans 6,577, HIndex:77).
<http://dx.doi.org/10.1016/j.rser.2013.07.058>
- C. Voyant, P. Randimbivololona, M-L. Nivet, **C. Paoli**, M. Muselli, « 24-hours ahead global irradiation forecasting using Multi-Layer Perceptron », *Meteorological Applications*, 2013 (IF2012 1,318 HIndex:26).
<http://10.1002/met.1387>.
- C. Voyant, M. Muselli, **C. Paoli**, M-L. Nivet, « Hybrid methodology for hourly global radiation forecasting in Mediterranean area », *Renewable Energy*, Volume 53, May 2013, Pages 1-11 (IF2011 2,978; IF5 ans 3,200, HIndex:60).
<http://dx.doi.org/10.1016/j.renene.2012.10.049>
- G. Notton, **C. Paoli**, L. Ivanova, S. Vasileva, M-L. Nivet, « Neural network approach to estimate 10-min solar global irradiation values on tilted planes », *Renewable Energy*, Volume 50, Pages 576-584, February 2013 (IF2011 2,978; IF5 ans 3,200, HIndex:60).
<http://dx.doi.org/10.1016/j.renene.2012.07.035>
- G. Notton, **C. Paoli**, M-L. Nivet, S. Vasileva, J-L Canaletti, C. Cristofari, « Estimation of hourly solar global irradiances on tilted planes from horizontal ones using artificial neural », *Energy*, Volume 39, Issue 1, Pages 166-179, March 2012 (IF2011 3,487; IF5 ans 3,858, HIndex:60).
<http://dx.doi.org/10.1016/j.energy.2012.01.038>

- C. Voyant, M. Muselli, **C. Paoli**, M-L. Nivet, « Numerical weather prediction (NWP) and hybrid ARMA/ANN model to predict global radiation », Energy, Volume 39, Issue 1, Pages 341-355, March 2012 (IF2011 3,487; IF5 ans 3,858, HIndex:60).
<http://arxiv.org/abs/1201.1613>
- C. Voyant, M. Muselli, **C. Paoli**, M-L. Nivet, « Optimization of an artificial neural network dedicated to the multivariate forecasting of daily global radiation », Energy, Volume 36, Issue 1, Pages 348-359, January 2011 (IF2011 3,487; IF5 ans 3,858, HIndex:60).
<http://dx.doi.org/10.1016/j.energy.2010.10.032>
- **C. Paoli**, C. Voyant, M. Muselli, M-L. Nivet, « Forecasting of preprocessed daily solar radiation time series using neural networks », Solar Energy, vol. 84, n°. 12, p. 2146-2160, December 2010 (IF2011 2,475; IF5 ans 2,902, HIndex:64).
<http://dx.doi.org/10.1016/j.solener.2010.08.011>

B. Publications dans d'autres revues internationales à comité de lecture

- C. Voyant, G. Notton, **C. Paoli**, M-L. Nivet, M. Muselli, K Dahmani, « Numerical weather prediction or stochastic modeling: an objective criterion of choice for the global radiation forecasting », International Journal of Energy Technology and Policy, 2013 (HIndex:8).
- G. Notton, **C. Paoli**, S. Vasileva, L. Ivanova, M-L. Nivet, L. Stoyanov, « Transformation of horizontal global solar data into tilted irradiation at hourly scale: comparison of various methodologies », Ecological Engineering and Environmental Protection, Vol. 3, Pages 13-19, 2012.

C. Publications dans des revues nationales à comité de lecture

- I.Caluianu, G.Notton, I.Colda, **C.Paoli**, Maximum power point prediction of a PV module using artificial neural networks. Sesiunea de comunicari stiintifice a scolii doctorale din universitatea tehnica de constructii bucuresti 23, pp. 51-58, 23ISBN: 978-973-100-129-6, editure conspress, 2010.
- I. Caluianu, G. Notton, I. Colda, **C. Paoli**, « Photovoltaic Module Maximum Power Point using One Diode Model and an Artificial Neural Network Model », Mathematical Modelling in Civil Engineering, n°1-2, pp. 33-38. ISSN 2066-6926, Mars 2011.

D. Participation à des ouvrages collectifs

- P. Oberti, **C. Paoli**, « Participative and multicriteria localization of wind farm projects in Corsica island: decision aid process and results », Springer book on MCDA applications. Ouvrage collectif, sous presse 2013.

- G. Notton, **C. Paoli**, M-L. Nivet, C. Voyant, « Prédire le vent et le soleil en s'inspirant du fonctionnement du cerveau », numéro spécial Energie, Stantari, sous presse 2013.
- L. Capocchi, **C. Paoli**, « Flash sous Linux avec Eclipse », numéro spécial Flash et Eclipse, phpSolutions, mai 2008.
- **C. Paoli**, « C2I - Mener des projets en travail collaboratif à distance (B7) » Ouvrage électronique sous la licence Creative Commons, version 1.0 & 2.0, C2I Mutualisé pour l'Enseignement Supérieur : <http://www.c2imes.org>, Indice Dewey: 005, 2005 & 2006.

E. *Communications avec actes*

- C. Voyant, M-L. Nivet, **C. Paoli**, M. Muselli, G. Notton, « Meteorological time series forecasting based on MLP modelling using heterogeneous transfer functions », International Conference on Mathematical Modeling in Physical Sciences 2014, Madrid : Spain (2014).
- G. Notton, **C. Paoli**, S. Diaf, « Estimation of Tilted Solar Irradiation Using Artificial Neural Networks », Mediterranean Green Energy Forum 2013: Proceedings of an International Conference MGEF-13, Volume 42, Pages 33–42, 2013.
<http://dx.doi.org/10.1016/j.egypro.2013.11.003>
- P. Haurant, C. Voyant, M. Muselli, M-L. Nivet, **C. Paoli**, « Hourly global radiation prediction from geostationary satellite data », International Conférence EU PVSEC, 2013.
- C. Voyant, D. Kahina, G. Notton, **C. Paoli**, M-L. Nivet, M. Muselli, P Haurant, « The global radiation forecasting based on NWP or stochastic modeling: an objective criterion of choice », International Conference ICNCRE, 2013.
- K. Dahmani, G. Notton, R. Dizene, **C. Paoli**, C. Voyant, M-L. Nivet, K Karouk, « Estimation of 5-min solar global irradiation on tilted planes by ANN method in Bouzareah », International Conference ICNCRE, 2013.
- C. Voyant, W. Tamas, **C. Paoli**, A. Balu, M. Muselli, M-L. Nivet, G. Notton, « Time series modeling with pruned multi-layer perceptron and 2-stage damped least-squares method », International Conference IC-MSQUARE, 2013.
- **C. Paoli**, C. Voyant, M. Muselli, M-L. Nivet, « Multi-horizon Irradiation Forecasting Using Time Series Models » International Conference ISES Solar World Congress, 2013.
- W. Tamas, G. Notton, **C. Paoli**, C. Voyant, M-L. Nivet, A Balu, « Urban ozone concentration forecasting with artificial neural network in Corsica », International Conference EENVIRO, 2013.
- W. Tamas, G. Notton, **C. Paoli**, J-L. Savelli, M-L. Nivet, C. Voyant, A. Balu, « Atmospheric pollutants concentration forecasting in Corsica », 3rd Charmex International Workshop. 5-9 Novembre 2012, Cargèse, France.
<http://charmex.lsce.ipsl.fr>.

- Z. Palková, G. Notton, **C. Paoli**, I. Okenka, « Web 2.0 technologies as a support for vocational education and training in the field of renewable energy resources », Colloque Francophone sur l'Energie, l'Environnement, l'économie et la Thermodynamique, COFRET2012, pp. 337-340, 11-13, Sozopol, Bulgarie, Juin 2012.
- G. Notton, **C. Paoli**, S. Vasileva, L. Ivanova, M-L. Nivet, L. Stoyanov, « Transformation of horizontal global solar data into tilted irradiation at hourly scale: comparison of various methodologies », Colloque Francophone sur l'Energie, l'Environnement, l'économie et la Thermodynamique, COFRET2012, pp. 260-265, Sozopol, Bulgarie, 11-13 juin 2012.
- Lambert et al. « Mediterranean atmospheric observatory in Corsica within the framework of HyMEX and ChArMEX », 5th HYMEX Workshop, Punta Prima, Sant Luis, Menorca, Spain, 17-19 May 2011.
- **C. Paoli**, G. Notton, M-L. Nivet, M. Padovani, J-L. Savelli, « Neural Network Model Forecasting for Prediction of Hourly Ozone Concentration in Corsica », 10th IEEE International Conference on Environment and Electrical Engineering, Roma, Italy, 2011.
- Lambert et al. « A Mediterranean atmospheric observatory in Corsica within the framework of HyMEX and ChArMEX », 4th HyMeX workshop at CNR, Bologna, Italy. 8-10 June 2010.
- C. Voyant, M. Muselli, **C. Paoli**, M-L. Nivet, « Optimization of an artificial neural network (ANN) dedicated to the daily global radiation and PV plant production forecasting using exogenous data », 25th European Photovoltaic Solar Energy Conference, Valence: Spain, 2010.
- **C. Paoli**, C. Voyant, M. Muselli, M-L. Nivet, « Use of exogenous data to improve an artificial neural networks dedicated to daily global radiation forecasting », 9th IEEE International Conference on Environment and Electrical Engineering, Prague, Czech Republic, May 2010.
- **C. Paoli**, C. Voyant, M. Muselli, M-L. Nivet, « Solar Radiation Forecasting Using Ad-Hoc Time Series Preprocessing and Neural Network », 5th IEEE International Conference on Intelligent Computing (ICIC) Ulsan: Korea, p. 898-907, 2009.
- F. Dulac et al. « Project ChArMEX – the Chemistry-Aerosol Mediterranean Experiment », SOLAS Open Science Conference, Barcelona, Spain, 16-19 November 2009.
- C. Voyant, M. Muselli, **C. Paoli**, M-L. Nivet, P. Poggi, « Predictability of PV power grid performance on insular sites without weather stations: use of artificial neural networks », 24th European Photovoltaic Solar Energy Conference, Hamburg: Germany, 2009.
- **C. Paoli**, M-L. Nivet, P. Oberti, « Semanticize data for decision support: development and deployment of an ontological vocabulary on the web of data », 8th Decision Deck Workshop, Corte : France, 2011.
- P. Oberti, **C. Paoli**, « A new support for participatory multicriteria decision aiding: Ev@I software », 73rd Meeting of the European Working Group « Multiple Criteria Decision Aiding », France (2011).

- **C. Paoli**, P. Oberti, M-L. Nivet, « Système d'information sémantique et géographique pour l'AMCD : état de l'art et construction d'une interface utilisateur », 72ème journées du groupe de travail européen « Aide multicritère à la décision », Ecole Centrale Paris, 7-9 octobre 2010.
- M-L. Nivet, **C. Paoli**, P. Oberti, « Use of the XMCD format to describe a multicriteria evaluation process to localize a wind park in Corsica: feedback », 5th Decision Deck Workshop, Telecom Bretagne, Brest, France, 17-18 September 2009.
- **C. Paoli**, P. Oberti, « Actors roles and software interface users rights in a multicriteria evaluation process: state of the art and proposals », 23rd European conference on operational research, Bonn: Germany, 2009.
- P. Oberti, **C. Paoli**, M. Ferracci, « Energies renouvelables, territoires d'accueil et participation : une expérimentation d'outils visant la médiation des préoccupations d'acteurs en région corse, Conférence », « Les outils pour décider ensemble Nouveaux territoires, nouveaux paradigmes », Université Paris Dauphine, Paris, France, 02 et 03 novembre 2006.
- **C. Paoli**, F. Etti, R. Papi, « 24 heures pour innover : sensibiliser des étudiants aux démarches d'innovation et à la création d'entreprise », 2èmes Rencontres des Pratiques Pédagogiques en Entrepreneuriat autour de la Méditerranée, 25 avril 2006, Tunis, Tunisie.
- **C. Paoli**, M-L Nivet, F. Bernardi, L. Capocchi, « Simulation-Based Validation of VHDL Description Using Constraints Logic Programming », IEEE International 5th Workshop on RTL and High Level Testing, Osaka, Japan, 11-12 November 2004.
- **C. Paoli**, M-L. Nivet, J-F. Santucci, T. Campana, « Path-Oriented Test Data Generation of Behavioral VHDL Description », IEEE International Workshop on Electronic Design, Test & Applications (DELTA'02), Christchurch, New Zealand, pp. 382-386, 29-31 November 2002.
- **C. Paoli**, M-L. Nivet, J-F. Santucci, « Use of constraint solving in order to generate test vectors for behavioral validation, communication », IEEE International High Level Design Validation and Test Workshop (HLDVT'00), Berkeley, California, USA, pp. 15-20, 8-10 November 2000.
- **C. Paoli**, M-L. Nivet, J.F. Santucci, « Test Vectors Generation using Path Selection and Constraint Solving for the Validation of Behavioral VHDL Design », 4th IEEE International Workshop on System Test and Diagnosis workshop (IWSTD'00), Atlantic City, New Jersey, USA, 5 p, 5-6 October 2000.
- **C. Paoli**, P-A. Bisgambiglia, D. Federici, J-F. Santucci, « High Level Validation of VHDL Description using Behavioral Fault Simulation », poster, actes du congrès international Environnement et Identité en Méditerranée, programme de coopération scientifique Interreg II, 13-16 juin, Corte, 2000.
- J-F. Santucci, **C. Paoli**, « High level test bench generation using software engineering concepts », 30th IEEE International Test Conference (ITC'99), Atlantic City, New Jersey, USA, 28-30 September 1999.

- **C. Paoli**, J-F. Santucci, « Validation of Behavioral VHDL Descriptions Using Software Engineering Concepts », 2nd IEEE Electronic Circuits and Systems Conference (ECS'99), Bratislava, Slovakia, pp. 215-218, 6-8 September 1999.
- **C. Paoli**, J-F. Santucci, « High Level Test Benches Generation for the Validation of VHDL Models of Microelectronic Systems », communication, 3rd IEEE International Workshop on System Test and Diagnosis workshop (IWSTD'99), Atlantic City, New Jersey, USA, 4 p, 30 September - 1 October 1999.

F. *Rapports divers*

- T. Antoine-Santoni, **C. Paoli**. Rapport interne, « Le C2i à l'Université de Corse - Pasquale PAOLI : Harmonisation, organisation du niveau 1 et projet de niveau 2 », 2010.
- **C. Paoli**, Rapport interne, « Dématérialisation des préinscriptions : version 2.0 », Université de Corse, France, 2008 :
- **C. Paoli**, M-L. Nivet, P. Oberti, « Application web pour la mise en œuvre de démarches d'AMCDP », Corte, France, 2006.
- **C. Paoli**, M-L. Nivet et J-F. Santucci, T. Campana, « Definition of the algorithm allowing paths conditions to be solved », Université de Corse, 11 p, decembre 2001.
- **C. Paoli**, M-L. Nivet et J-F. Santucci, T. Campana, « Validation of the Software S1 described in the Technical Report N°4 on a set of benchmark VHDL descriptions », Université de Corse, 33 p, juin 2001.
- **C. Paoli**, M-L. Nivet et J-F. Santucci, T. Campana, « Validation of the Software S2 described in the Technical Report N°5 », Université de Corse, 11 p, juin 2001.
- M-L. Nivet, **C. Paoli** et J-F. Santucci, « Definition of the internal model based on graph structures representing VHDL behavioral descriptions », Université de Corse, 11 p, septembre 2000.
- J-F. Santucci, **C. Paoli** et M-L. Nivet, « Algorithms allowing to automatically generate an internal model from a VHDL behavioral description », Université de Corse, 13 p, septembre 2000.
- **C. Paoli**, M-L. Nivet et J-F. Santucci, « Definition of the algorithm allwing the minimum set of path to be generated », Université de Corse, 10 p, septembre 2000.
- **C. Paoli**, M-L. Nivet et J-F. Santucci, « Software Implementation of the automatic generation of the internal model », Université de Corse, 11 p, septembre 2000.
- **C. Paoli**, M-L. Nivet et J-F. Santucci, « Software implementation of the minimum set of paths generation - S2 », Université de Corse, 10 p, septembre 2000.
- R. Kozakowski, **C. Paoli**, D. Saarinen, « Overview of Test Methods Using Boundary Scan », University of New Hampshire, Durham, NH, USA, 28 p, decembre 1999.

VII. Activités de recherche

A. Avant-propos

La suite de ce document propose de détailler les activités de recherches développées depuis mon recrutement en septembre 2001 en tant que Maître de Conférences à l'Université de Corse. Rattaché au laboratoire « Systèmes Physiques pour l'Environnement », devenu par la suite « Sciences Pour l'Environnement » (SPE) et qui est une unité mixte de recherches CNRS sous l'intitulé UMR 6134, mes activités de recherche peuvent être déclinées en deux grandes périodes.

La première période correspond à mes activités de recherche débutées au sein de l'équipe "Modélisation et Simulation" et concerne l'application de techniques de test de logiciel pour des descriptions de circuit électroniques décrits avec le langage VHDL. Il s'agit d'un langage de programmation utilisé dans l'industrie et permettant de modéliser et simuler des circuits électroniques mais aussi de les synthétiser, c'est à dire obtenir une description en terme de portes logiques de ceux-ci en vue de leur fabrication. Une fois synthétisé, il devient impossible de tester de façon exhaustive un circuit électronique en un temps raisonnable. En effet, le circuit électronique alors constitué de millions de portes logiques, contient un nombre de chemins à tester et donc à parcourir trop important pour tous les parcourir. Mon travail de recherche s'est ainsi focalisé sur la génération automatique de jeux de test à partir de descriptions écrites en VHDL. Ce type de description étant similaire à un programme, nous avons exploré des techniques utilisées dans le domaine du test de logiciels.

La seconde période fait suite à un changement d'équipe opéré en 2005. En effet, j'ai eu l'opportunité de m'investir dans deux thématiques propres à l'étude des systèmes énergétiques à sources d'Energies Renouvelables (EnR) et leur intégration dans les réseaux électriques insulaires, comme celui de la Corse. Au sein de l'équipe EnR de l'UMR 6134 j'ai pu développer des travaux de recherche au niveau de deux axes principaux.

Le premier concerne la modélisation et la prédiction de phénomènes spatio-temporels. Nous nous sommes plus particulièrement intéressés à l'utilisation des réseaux de neurones artificiels appartenant aux techniques d'intelligence artificielle. Nous les avons principalement utilisées dans le cadre de la prédiction de séries temporelles d'ensoleillement et de pics de pollution atmosphérique. Nous avons étudié tout d'abord l'impact de prétraitements statistiques des séries temporelles sur la performance de prédicteurs. Aussi, au vue du nombre important de données dont nous disposions, nous avons étudié différentes méthodes de sélection de caractéristiques (ou de variables) avec l'objectif d'obtenir le modèle le plus performant mais aussi le plus simple possible (principe de parcimonie). Enfin, lors de la comparaison de performance entre notre modèle et différentes méthodes de prédictions ("benchmarking" en anglais), nous avons constaté un chevauchement de

celles-ci suivant l'horizon considéré. Nous avons alors étudié comment les hybrider, technique qui entre dans le cadre de la problématique plus large de la sélection de modèles. Une étude sur l'extension de ces méthodes pour les séries temporelles de vitesse de vent a été initiée et sera évoquée en perspectives dans la conclusion de ce mémoire. Enfin et plus récemment nous avons démarré des travaux concernant l'estimation de potentiel solaire également à l'aide réseaux de neurones artificiels. Il s'agit cette fois, non pas de prédire, mais d'estimer le rayonnement solaire incliné à partir du rayonnement solaire global horizontal et d'autres variables météorologiques disponibles.

Le second axe propose d'étudier l'apport des langages informatiques de représentation des connaissances pour représenter la connaissance produite tout au long des processus d'aide à la décision. Des projets tels que les centrales photovoltaïques ou encore les fermes éoliennes constituent des exemples significatifs de processus de décision extrêmement complexe, difficile et conflictuel.

L'ensemble de ces travaux de recherche seront décrits dans trois sous-parties distinctes. La prédiction et modélisation de phénomènes spatio-temporels sera traitée en premier car il s'agit de la partie la plus volumineuse tant en terme de production que de collaborations scientifiques. La thématique de l'aide à la décision, plus récente et plus prospective sera traitée ensuite. Enfin, les travaux plus anciens et mis en veille concernant l'utilisation de techniques de test de logiciel pour des circuits électroniques décrits en VHDL concluront cette synthèse de mes travaux de recherche.

Mais avant de détailler ces trois axes de travail, il m'a semblé intéressant de situer chacun d'eux par rapport au domaine de l'informatique théorique. Notamment dans le cadre de sous-disciplines tel que la théorie de la calculabilité, de la complexité et de l'algorithmique, qui propose de comprendre les limites de l'informatique et notamment de distinguer quels sont les problèmes solubles et insolubles par des algorithmes. On parle respectivement de problèmes indécidables et de problèmes décidables.

B. Décidabilité et calculabilité

L'informatique ne peut pas tout et il a été démontré dans le cadre de l'informatique théorique [1] que pour certain problème il n'existait pas d'algorithme. Ces problèmes pour lesquels il n'existe aucune méthode de résolution sont appelés problèmes indécidables. Cette notion a été formalisée il y a plus d'un demi-siècle par le logicien Gödel [2]. Le problème de l'arrêt ("halting problem" en anglais) est un exemple très connu de problème indécidable. En effet, il n'existe pas d'algorithme qui permette de s'assurer qu'un algorithme ne bouclera pas à l'infini. Cependant même si un problème est indécidable et qu'il ne faut donc pas perdre son temps à en chercher une solution, on peut chercher un problème dérivé mais plus simple.

La prévision météorologique est typiquement un problème indécidable. On peut trouver dans [3] une démonstration qui propose de réduire le problème de la prévision au problème de l'arrêt. Le caractère indécidable de la prévision est également évoqué dans [4]. Dans [5], la prévision météorologique est comparée à un autre problème indécidable, celui du jeu de la vie [6]. Il s'agit en fait d'un automate cellulaire qui fonctionne sur une grille à cases carrées (les cellules) dont l'ensemble constitue l'espace plan où tout se déroule. Chaque cellule est vide ou occupée, morte ou vivante. Le temps s'y écoule de manière discrète : il y a un instant 0, un instant 1, un instant 2, etc. Chaque cellule n'a connaissance que de l'état de ses huit voisines directes, et la règle qui détermine l'évolution des cellules d'une génération à la suivante est la règle suivante : naissance si trois voisins, survie si deux ou trois voisins. Malgré cet univers extrêmement simple, la non-prévisibilité du système a été démontrée.

Comme d'autres problèmes indécidables, le problème de la prévision météorologique a lui aussi été restreint et résolu par différentes approches que l'on classe généralement en deux grands groupes : le groupe des modèles de prévisions numériques de temps et le groupe des modèles basés sur des approches statistiques. Ces méthodes font l'objet d'une présentation dans le paragraphe suivant.

Mais l'étude théorique ne s'arrête pas là, on doit ensuite déterminer la calculabilité. En effet, savoir qu'un problème est décidable, c'est-à-dire qu'il existe des algorithmes qui le résolvent, ne suffit pas. Il faut aussi que l'algorithme soit efficace, c'est-à-dire en utilisant des ressources raisonnables, en temps de calcul et en espace mémoire. La théorie de la complexité permet ainsi de définir un critère d'efficacité qui ne dépende pas de la machine sur laquelle sera exécuté l'algorithme, ou du langage de programmation dans lequel il sera écrit. Des problèmes bien connus dans le domaine du test de logiciel ou de l'aide à la décision sont décidables mais avec des complexités trop importantes pour être exécutés en un temps raisonnable. Nous y reviendrons dans les paragraphes dédiés à ces thématiques.

C. *Prédiction et modélisation de phénomènes spatio-temporels*

1. Introduction

La modélisation et la prédiction de phénomènes spatio-temporels sont des préoccupations majeures dans nos sociétés, et tout particulièrement dans le domaine des sciences environnementales, domaine dans lequel s'inscrivent les projets de recherche structurant le laboratoire Sciences Pour l'Environnement (SPE) UMR CNRS 6134. Dans ce contexte, l'équipe Energies Renouvelables (EnR) de l'UMR 6134 s'intéresse à la prédiction et la modélisation de séries temporelles de données environnementales mesurées.

En effet, l'un des éléments clés pour un pilotage optimal de systèmes à source renouvelable d'énergie, consiste en la caractérisation de la ressource (soleil ou vent) afin de pouvoir mieux la contrôler selon les besoins. Il peut être ainsi nécessaire selon les opportunités de stocker ou de déstocker l'énergie. L'insertion des EnR sur les réseaux îliens non interconnectés est un enjeu majeur pour accroître, d'une part, leurs indépendances énergétiques et d'autre part, réduire les factures énergétiques et environnementales. Cependant, la sûreté d'un réseau électrique insulaire peut être mise à mal lorsque la proportion d'EnR intermittente injectée sur le réseau dépasse un certain seuil. Leur caractère fatal empêche le producteur de piloter sa production et ainsi handicape leur intégration dans les réseaux électriques. Un seuil réglementaire de 30% a ainsi été fixé par décret.

Une autre problématique intéressante concerne plus particulièrement la gestion optimale des équipements de production d'énergie solaire. L'équipe EnR a souhaité ainsi investiguer l'interaction entre aérosols et rayonnement solaire. En effet, il existe actuellement très peu d'études sur les impacts des pollutions sur le rayonnement solaire et donc leurs conséquences sur la productivité des systèmes énergétiques solaires.

Enfin, plus récemment encore, une partie de l'équipe EnR s'est intéressée à l'utilisation des réseaux de neurones artificiels non plus pour prédire mais pour estimer le rayonnement solaire incliné à partir du rayonnement solaire global horizontal et d'autres variables météorologiques disponibles.

Pour répondre à ces problématiques, il est indispensable de mettre en place des outils et des méthodes de modélisation et de prédiction. Le soleil, le vent, les aérosols, étant des phénomènes environnementaux spatio-temporels, étudier des méthodes de modélisation et prédiction pour ces ressources nous a conduits à nous intéresser à la météorologie et aux modèles de prévision météorologiques actuellement en usage. Parallèlement, disposant d'un historique de données sur ces ressources, nous nous sommes intéressés aux travaux effectués dans le domaine de



Université de Corse - Pasquale Paoli
Université Galatasaray
-
UMR CNRS 6134
Sciences Pour l'Environnement
Ministère des Affaires Etrangères



la prévision de modèles économiques, et notamment la prédiction de séries temporelles par des approches statistiques.

2. Les modèles de prédiction

Le nombre impressionnant d'articles publiés sur le sujet nous empêchant d'en proposer une liste exhaustive, nous nous limitons dans les deux prochaines sections aux principales méthodes appartenant à chacune des deux grandes familles de modèle de prédiction. Puis, après avoir présenté les façons de comparer en termes de performances ces modèles, nous proposerons une synthèse de leur utilisation suivant la résolution temporelle et/ou spatiale sur laquelle on se place.

a) Les modèles de prévision numérique du temps

Les modèles de prévision numérique du temps (NWP en anglais pour "Numerical Weather Prediction") sont des modèles qui s'appuient sur des considérations de la physique de l'atmosphère. A partir de l'état actuel de l'atmosphère, un système d'équations permet d'estimer l'évolution de certaines variables d'état telles que température, pression, humidité, vitesse du vent, etc. L'acquisition des données nécessaires est réalisée par une grande diversité de moyens techniques, fixes et mobiles mesurant les paramètres de l'atmosphère : stations météo, capteurs, radars, ballons, radiosondes, satellites, avions, bouées et navires. Les équations sont ensuite résolues par un ensemble de programmes informatiques qui simulent le mouvement atmosphérique dans l'espace et le temps. Une variété de phénomènes météorologiques peut être ainsi analysée et prédite par ces types de modèles de prévision numérique du temps.

Dans ce type de modèle, l'atmosphère est donc représentée par une grille 3D. Plus le maillage est fin, plus la simulation sera élaborée et précise. La simulation de ce modèle génère l'état futur de l'atmosphère dans chacun des points du réseau à partir de son état initial [7]. La figure ci-après illustre les précisions de maillage relatives à trois modèles différents utilisés par Météo-France¹ : ARPEGE, ALADIN et AROME. Météo-France est l'organisme français de météorologie (établissement public administratif) chargé de la prévision et de l'étude des phénomènes météorologiques.

¹ <http://www.meteofrance.fr/prevoir-le-temps/la-prevision-du-temps/les-modeles-de-prevision-de-meteo-france>

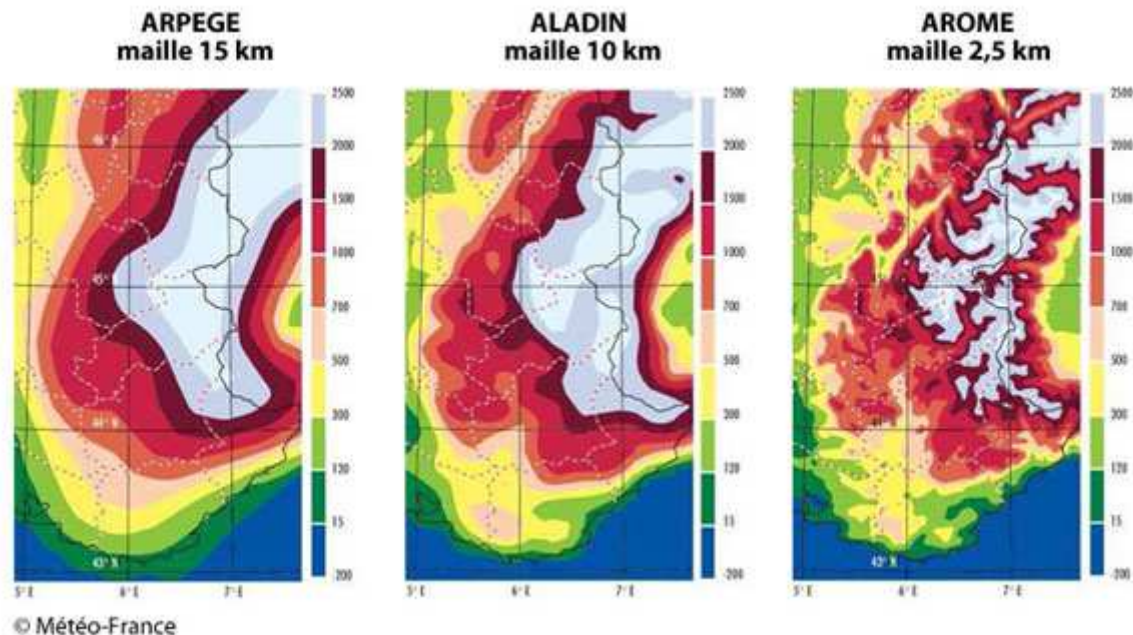


Figure 1. Exemple de maillage pour la région des Alpes pour des maillages différents.

En réalité un modèle NWP n'est pas un outil unique et monolithique mais plutôt un système de prévision numérique qui repose sur l'exploitation d'une chaîne de modèles complémentaires. On distingue en deux classes ces modèles. Les modèles globaux, comme ARPEGE ou encore le modèle du Centre européen de prévision météorologique à moyen terme (CEPMMT), couvrent l'ensemble du globe avec une maille moyenne de 25 km. Ils permettent de prévoir les phénomènes de grande échelle (dépressions, anticyclones par exemple) qui parcourent le globe. Les modèles régionaux ou méso-échelle, comme ALADIN, AROME ou encore le modèle "Weather Research Forecasting" (WRF) conçu par le "National Center for Atmospheric Research" (NCAR), dispose de résolution plus faible (entre 2,5 et 8 km). On notera également l'existence du modèle de chimie-transport CHIMERE dédié à la pollution. Ces modèles permettent de prévoir le temps sur un domaine limité jusqu'à deux jours d'échéance.

Il est également intéressant de noter l'importance du facteur humain dans la prévision météo. En effet, tournant sur un supercalculateur, les modèles NWP fourniront des dizaines de scénarios différents d'évolution du temps. Les prévisionnistes analyseront alors les scénarios et les ajusteront en fonction de leur connaissance du climat régional et des limites des modèles numériques. Cette approche ne permettant pas d'évaluer les incertitudes qui pèsent sur le scénario de prévisions retenu, les prévisionnistes compléteront leur analyse en utilisant une méthode appelée prévision d'ensemble (ou probabiliste). Elle consiste à réaliser des simulations à partir de plusieurs descriptions légèrement différentes, représentatives de l'état initial de l'atmosphère. Les prévisionnistes pourront ainsi choisir le scénario le plus probable et quantifier l'incertitude qui pèse sur cette prévision.

b) Les modèles statistiques

Ces approches consistent en général en la création de modèles basés sur l'analyse de séries temporelles. On distingue deux grandes catégories dans cette famille : les modèles linéaires et les modèles non linéaires.

1. Les modèles linéaires

- La persistance

Il s'agit d'un modèle de type "naïf" très souvent utilisé, notamment à des horizons courts (de quelques minutes à quelques heures). Il s'agit en fait de la répétition d'une mesure de l'instant t à l'instant $t+h$, avec h l'horizon considéré.

- Les modèles autorégressifs de type Box-Jenkins

Les modèles de Box-Jenkins [8][9] sont les principaux modèles de séries temporelles en statistique. Ils s'appuient sur les propriétés d'autocorrélation et le plus simple d'entre eux est le modèle autorégressif (AR). Un AR d'ordre " n " est un modèle selon lequel la valeur future est obtenue à un bruit près, comme une combinaison linéaire des " n " dernières valeurs mesurées.

En augmentant le degré de complexité, il vient ensuite les modèles autorégressifs à moyenne mobile appelés ARMA, pour "Auto Regressive Moving Average" en anglais.

L'ensemble des modèles autorégressifs nécessitant une hypothèse de stationnarité, il existe un modèle plus général nommé ARIMA où le "I" désigne "Integrated" en anglais. Les modèles ARIMA permettent ainsi de traiter les séries non stationnaires après avoir déterminé le niveau d'intégration.

Les modèles SARIMA, où le "S" désigne "Seasonnal" en anglais, sont une extension des modèles ARIMA. Ils permettent d'y inclure une dynamique plus réaliste, c'est à dire les comportements saisonniers.

Enfin, pour prendre en compte des variables de type exogènes, les modèles VARIMA et les ARIMAX ont été définis où respectivement le "V" et le "X" signifient "Vector" et "with eXplanatory variables".

Bien que possédant d'excellentes qualités prévisionnelles, les modèles de type Box-Jenkins souffrent d'un défaut majeur. En effet, ils ne permettent pas de capturer les propriétés de phénomènes non linéaires.

2. Les modèles non linéaires

D'autres techniques plus récentes, basées sur l'intelligence artificielle, ont été prises en considération dans le cadre de la prévision de phénomènes environnementaux.

En général, une quantité de données importante est nécessaire à la phase d'apprentissage automatique de ces modèles qui définissent une relation entre leurs entrées et leur sortie, i.e. la variable que l'on cherche à prédire à un horizon h . Ces modèles, dits de type boîtes noires, permettent de modéliser des processus non-linéaires contrairement aux méthodes linéaires classiques. Cette famille comprend les Réseaux de Neurones Artificiels (RNA) [10], [11] sur lesquels nous avons choisi de nous concentrer et plus précisément sur un type d'architecture. En effet, plusieurs types de réseaux de neurones existent, avec différents types de neurones et différentes connexions entre eux, chacun spécialisé pour une problématique : la classification, la reconnaissance de motifs, la modélisation de fonction ou encore la prédiction. Le Perceptron Multi-Couche (PMC) est dans la littérature l'architecture la plus utilisée et la plus simple à mettre en œuvre.

Les PMC sont constitués d'une interconnexion d'unités de calcul simples qu'on appelle neurones artificiels et dont le fonctionnement s'inspire des connaissances sur les neurones biologiques. Chaque neurone reçoit des données en entrée, les traite et produit une sortie qui est propagée vers les neurones suivants. Les données reçues par le neurone sont d'abord multipliées par un poids propre, puis la somme de ces entrées pondérées, additionnée d'un biais propre au neurone devient l'argument de la fonction d'activation du neurone. Le résultat représente la sortie du neurone.

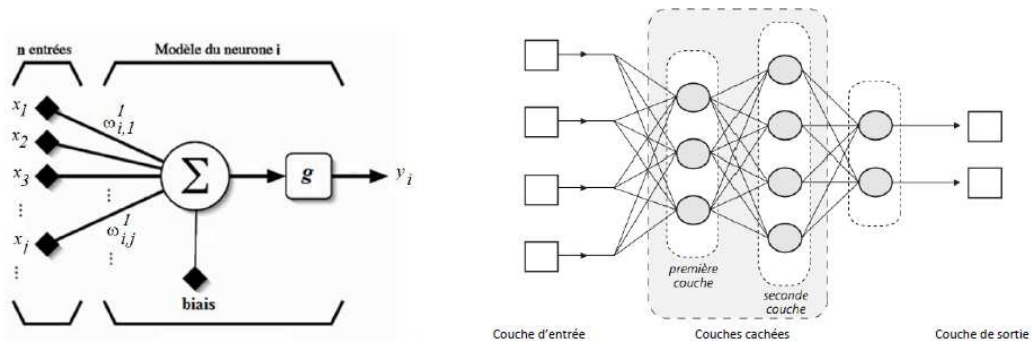


Figure 2. Neurone formel et architecture d'un PMC.

Les derniers neurones du réseau produisent les sorties du modèle. Les neurones qui les précèdent sont appelés les neurones de la couche cachée ou "hidden layers" en anglais.

Un PMC correspond donc à une fonction paramétrique à plusieurs variables, et ces paramètres sont les poids et les biais des neurones. L'apprentissage automatique permet de fixer les valeurs de ces paramètres, à l'aide de données qui seront utilisées par un algorithme d'apprentissage qui peut être supervisé ou non-supervisé. Dans le cas d'un apprentissage supervisé, l'algorithme optimise les paramètres pour faire tendre la sortie du réseau vers la valeur des données cibles fournies par

l'utilisateur. Un apprentissage non-supervisé n'a pas de cible et l'algorithme optimise les poids et biais uniquement en fonction des données en entrée.

Outre le fait que le PMC a été très largement utilisé dans de nombreux domaines, il a été démontré que ce type de réseau (une couche de neurones cachés en nombre fini, possédant tous la même fonction d'activation, et un neurone de sortie linéaire) pouvait modéliser n'importe quelle fonction indéfiniment dérivable (fonction lisse). Il est ainsi qualifié d'approximateur universel [12].

On peut également citer la possibilité d'agencer les réseaux "en cascade". Le modèle est alors constitué de plusieurs PMC, chacun effectuant une prédiction à un horizon spécifique et pouvant donc bénéficier d'entrées distinctes. Leur sortie est ensuite réutilisée par le PMC suivant qui réalise une prédiction à un horizon supérieur.

Dans le cadre de nos travaux nous avons utilisé principalement l'architecture PMC entraînés à l'aide de l'algorithme de Levenberg-Marquardt (LM). La figure suivante illustre l'utilisation d'un PMC dans le cas de la prédiction de séries temporelles.

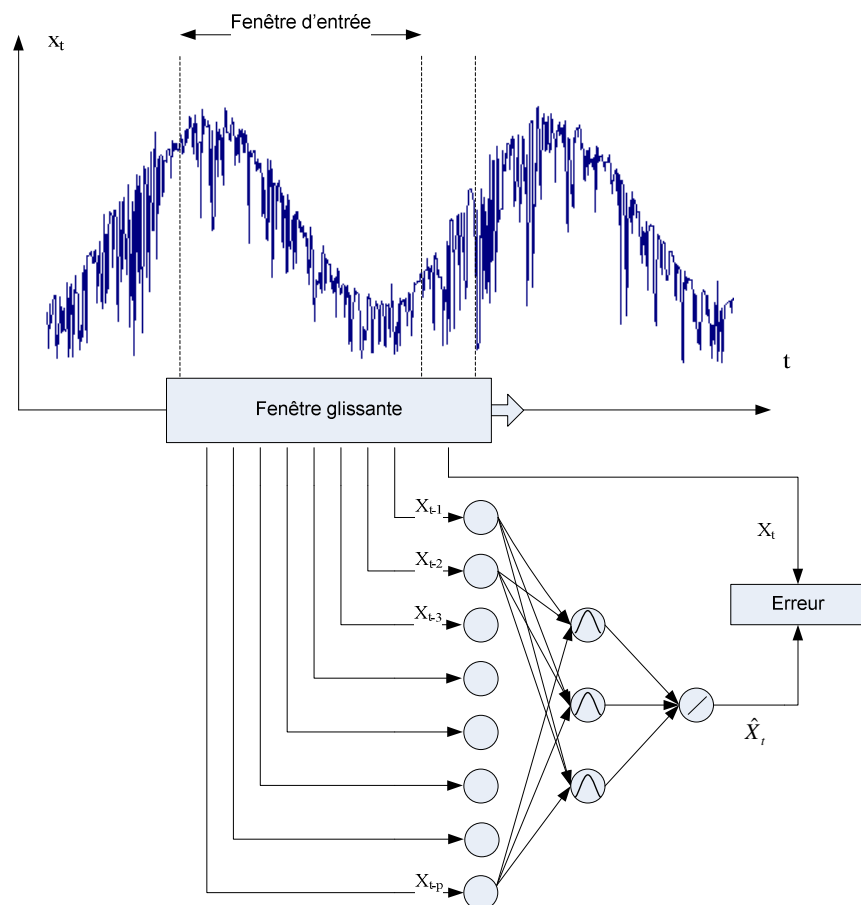


Figure 3. Construction d'un PMC utilisable dans le cas de prédiction de séries temporelles.

c) Comparaison des performances

De nombreux critères existent pour vérifier si un modèle de prédiction est efficace. Parmi les plus classiques on citera : MAE (Mean Absolute Error), MBE (Mean Bias Error), RMSE (Root Mean Square Error), nRMSE (Normalized Root Mean Square Error), R^2 et Index of Agreement (IA).

Dans la littérature on retrouve souvent l'erreur nRMSE comme indicateur de référence. L'erreur MBE est représentative du comportement global d'un système puisqu'elle est une moyenne de l'erreur brute. Si le critère de performance retenu est le comportement global du système alors l'erreur MBE peut être adaptée. L'erreur MBE nous donne une idée de l'erreur en termes de sous ou sur estimation. Cependant l'erreur MBE autorise la compensation des erreurs. Pour caractériser le comportement local à long terme de la prédiction, il faut s'affranchir des phénomènes de compensation. Les erreurs MAE et RMSE sont construites sur ce principe. L'erreur MAE retranscrit le comportement local du système de manière linéaire, alors que l'erreur RMSE utilise l'erreur quadratique. Souvent décrit, le coefficient de détermination (R^2) est un indicateur qui permet de juger la qualité d'une régression linéaire, simple ou multiple. D'une valeur comprise entre 0 et 1, il mesure l'adéquation entre le modèle et les données observées. L'IA contrairement aux critères plus classiques basées sur des erreurs moyennes, représente le degré avec laquelle la prévision est sans erreur (0 pour le pire, 1 pour le mieux).

Dans le domaine des EnR, il est souvent très difficile de comparer les performances des méthodes de prédictions publiées par les chercheurs. En effet, les phénomènes environnementaux à prédire sont très spécifiques à la géographie (milieu insulaire, désertique, montagneux, etc.). Comparer des écarts de valeurs ou des pourcentages n'a souvent que peu d'intérêt. L'idéal serait de disposer de jeux de données et de méthodes au sein d'un "benchmark" à l'instar de ce qui se fait dans le domaine de prédiction de série temporelles en économétrie [13], [14].

Dans ce contexte, on peut signaler que l'action COST (European Cooperation in Science and Technology) WIRE² (Weather Intelligence for Renewable Energies) qui rassemble différents acteurs européens du domaine des énergies solaires et éoliennes (centres de recherche et industries) a initié un référencement des différentes méthodes de prédiction. Ce groupe rassemble différents acteurs européens du domaine des énergies solaires et éoliennes (centres de recherche et industries) dans le but d'améliorer les techniques de prévision météorologique dédiées à l'énergie.

² <http://www.wire1002.ch>

d) Différentes méthodes pour différentes échelles

Dans le cadre de la prévision de phénomènes spatio-temporels, les méthodes que nous venons de voir sont classifiables en terme de performance suivant la résolution temporelle et/ou spatiale sur laquelle on se place. La figure suivante extraite de [15] résume et illustre cette classification.

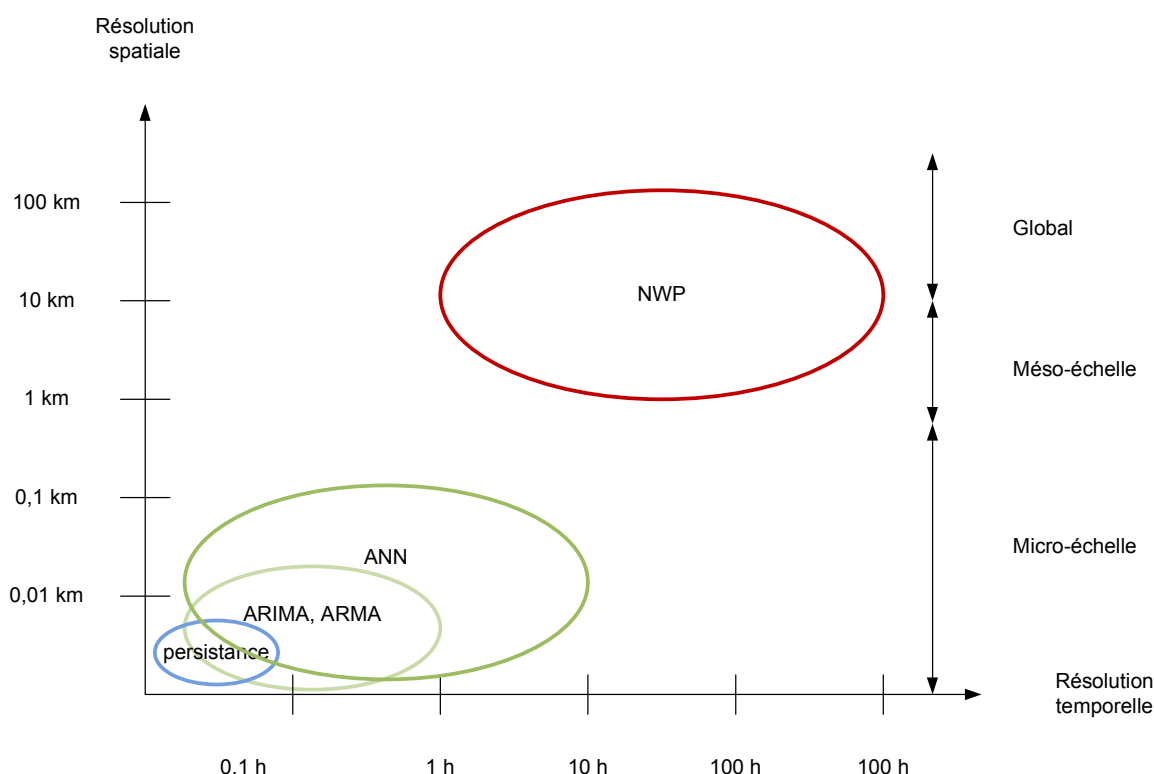


Figure 4. Classification des méthodes de prédiction en fonction de la résolution spatiale et temporelle.

Il est admis que les modèles NWP fournissent de meilleures prévisions que les méthodes purement statistiques (cf. paragraphes précédents) quand on considère des horizons allant de plusieurs heures à plusieurs jours. Cependant, la résolution spatiale des modèles NWP est souvent insuffisante pour modéliser les effets plus locaux comme les milieux insulaires ou montagneux auquel nous nous sommes intéressés plus particulièrement.

En règle générale, l'utilisation de méthodes statistiques est pertinente si l'on s'intéresse à des horizons de prévision courts (de l'ordre de la minute à plusieurs heures). La prévision à de tels horizons trouve un intérêt particulier dans le cas de sites isolés comme dans les îles par exemple.

La prédiction à différents horizons et pas de temps des séries temporelles est un problème qui recouvre de nombreux domaines d'application. Les études menées dans la finance et l'économétrie ont permis de dégager de nombreux modèles. Ces derniers ont été repris dans d'autres problématiques, dont la prédiction de données mesurées en lien avec l'environnement (ensoleillement, vent, pollution, etc.) à laquelle nous nous intéressons.

Nous nous sommes plus particulièrement attachés à l'étude de l'architecture MLP entraînés à l'aide de l'algorithme de Levenberg-Marquardt (LM) sur des séries temporelles.

3. Problématiques relatives aux méthodes statistiques

Dans le cadre de l'utilisation des méthodes statistiques plusieurs difficultés sont à résoudre. Nous devons tout d'abord disposer d'un historique de données suffisant et de qualité. Une analyse fine et des prétraitements sont nécessaires avant la paramétrisation du modèle. Un second problème important est la sélection des variables d'entrée du modèle. En effet, si l'on veut disposer du modèle le plus parcimonieux possible, il nous faut définir des méthodes et des critères permettant de sélectionner les bonnes variables en entrée de notre modèle. Enfin, un dernier problème auquel on peut être confronté est de vouloir utiliser une combinaison de plusieurs modèles afin d'augmenter la performance de la prédiction. Ces trois problématiques sont décrites plus précisément dans les sections suivantes.

a) Prétraitement des données

En général les méthodes de prédictions statistiques se révèlent plus performantes si les données subissent en amont certains traitements. Plusieurs traitements de données sont possibles et on peut citer les suivants comme les plus importants :

- Remplacement des données manquantes ou aberrantes : la plupart des jeux de données comportent des données manquantes, qui peuvent être dues aux périodes de maintenance des appareils ou à des dysfonctionnements ponctuels dans la chaîne d'acquisition. Il est donc important de connaître leurs répartitions, leurs fréquences et d'étudier le meilleur moyen de les traiter. Ces valeurs manquantes sont problématiques et ne peuvent pas être traitées par le modèle pour lequel des opérations mathématiques sont définies pour des valeurs réellement mesurées. Plusieurs méthodes de traitement des valeurs manquantes existent, une description et évaluation de celles-ci est proposée dans [14].
- Normalisation des données : du fait que certaines variables peuvent avoir plus d'influence que d'autres de par la différence d'amplitude que peuvent prendre certaines valeurs, il est nécessaire de les normaliser.
- Analyse statistique : avant de paramétrer tout modèle de prédiction, il est nécessaire d'analyser les données à disposition, étudier la façon dont elles évoluent au cours du temps et les différents liens entre les variables. Pour cela on doit procéder à une étude des autocorrélations, des spectres, des tendances, etc. Cette étape est un préalable au prochain prétraitement évoqué.
- Stationnarisation des variables : ce prétraitement tient une place importante dans l'analyse et la prévision des séries temporelles car la plupart des modèles de prévision sont structurés pour faire face à des séries

stationnaires, ou au moins faiblement stationnaires. Il existe plusieurs façons de traiter des séries temporelles non stationnaires et les transformer en une forme appropriée [16].

b) Sélection de caractéristiques

La sélection de caractéristiques [17] ou de variables ("feature selection" en anglais) est un problème difficile et constitue à l'origine un domaine très actif dans le domaine de la fouille de données ("data mining" en anglais). En effet, la fouille de données dans de très grandes bases devient un enjeu crucial pour des applications telles que le génie génétique, la finance, les études de marché, les processus complexes. L'explosion des volumes de données de nature informatique à traiter, on parle de « Big Data », a d'ailleurs permis de remettre en lumière ce type de travaux.

Dans le domaine de la modélisation de séries temporelles, la sélection de caractéristiques a également fait l'objet de nombreuses recherches.

Les intérêts de la sélection de variables peuvent se résumer comme suit :

- Si le nombre de variables est vraiment trop important, l'algorithme d'apprentissage ne peut pas terminer en un temps convenable. La sélection réduit alors la dimension de l'espace des caractéristiques.
- La réduction de la dimension de l'espace des caractéristiques permet de réduire le nombre de paramètres nécessaires à la description de ce modèle, donc de le rendre plus simple.
- La sélection de variables améliore la performance du modèle, sa vitesse et son pouvoir de généralisation.
- La sélection de variable permet de mettre à jour les liens qui unissent les variables au phénomène étudié et ainsi de participer à la compréhension du processus sous-jacent.

Une structure générale des algorithmes de sélection de caractéristiques peut être proposée comme illustrée sur la figure suivante. Jusqu'à ce qu'un certain critère soit satisfait, des sous-ensembles sont générés, en parcourant l'espace des sous-ensembles, et évalués.

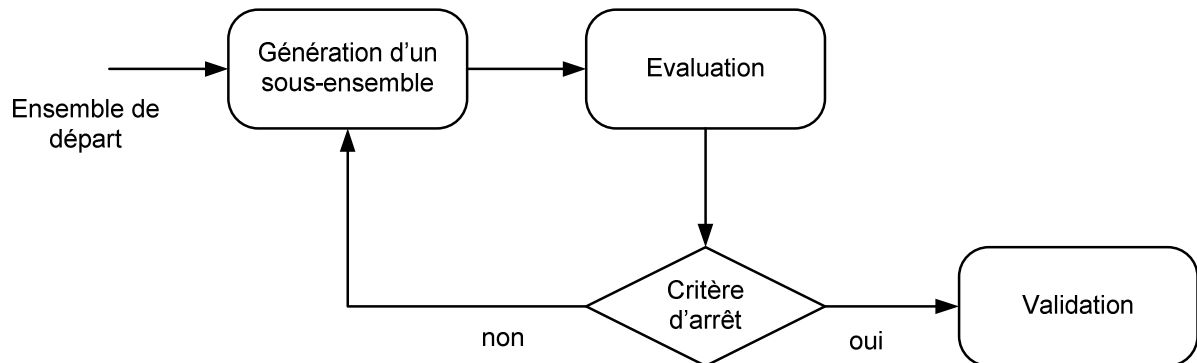


Figure 5. Procédure générale de sélection d'un sous-ensemble de caractéristiques.

L'évaluation d'un sous-ensemble peut être traitée de façons très diverses et il existe deux grandes classes de méthodes :

1. Les méthodes à adaptateurs ("wrappers" en anglais) qui utilisent le système pour évaluer la pertinence d'un groupe de variables ;
2. Les méthodes filtres ("filters" en anglais) qui sont indépendants du système, mais se basent sur des considérations statistiques, entropiques, de cohérence, de distance, etc.

Dans le cadre de nos travaux, nous nous sommes orientés vers les méthodes de type filtres qui sont plus rapides que les méthodes à adaptateurs. En effet, elles reposent sur des considérations plus théoriques et permettent de mieux comprendre les relations de dépendance entre variables.

c) Hybridation et sélection de modèles

Nous avons déjà pu constater que de nombreuses méthodes de prédiction ont été développées par des experts du monde entier et que selon l'horizon de prédiction (jours, heures, minutes) et la résolution spatiale considérée (1 à plusieurs kilomètres), on observe pour les différents modèles des performances très diverses qui peuvent se recouvrir.

Ainsi, lorsque que pour un problème il est difficile d'avoir un seul algorithme qui résout de manière optimale toutes ses instances, il est courant d'élaborer des approches combinant plusieurs algorithmes résolvant le même problème. Les approches qualifiées d'approches hybrides permettent la combinaison d'algorithmes pouvant être mise en œuvre au niveau système (en construisant des bibliothèques, des langages et composants adaptatifs etc.) ou au niveau purement algorithmique.

La technique d'hybridation peut en réalité être rattachée à une problématique plus large appelée sélection de modèles ou encore problème de la sélection d'algorithme. Il s'agit de trouver l'algorithme à exécuter qui maximise une performance, par exemple le temps d'exécution ou encore un taux d'erreur. Pour cela, on mesure la performance des algorithmes à l'aide de caractéristiques. À partir de ces données, on crée un modèle par régression qui prédit la performance de l'algorithme pour toute instance du problème.

La figure suivante propose une vue schématique du problème de la sélection d'algorithmes formalisé par Rice en 1976 [18].

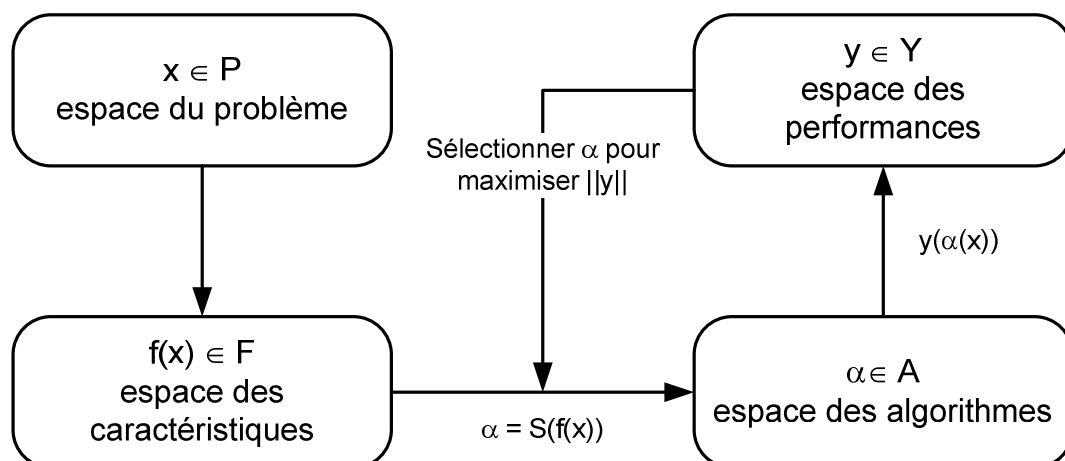


Figure 6. Vue schématique de la formalisation du problème de la sélection d'algorithmes.

Le modèle de la sélection d'algorithmes possède quatre composantes essentielles [19] :

- P , l'ensemble des instances du problème ;
- F , l'espace des caractéristiques d'instance ;
- A , l'ensemble des algorithmes de résolution ;
- Y , l'espace des performances qui représente les vecteurs des performances de chaque algorithme pour chaque instance.

La sélection d'algorithme repose sur une fonction d'application $S : F \rightarrow A$. Pour une instance donnée $x \in P$, avec des caractéristiques $f(x) \in F$, la fonction $S(f(x))$ retourne l'algorithme $\alpha \in A$ qui est le plus adapté pour résoudre x , i.e. tel que la mesure de performance $\|y(\alpha(x))\|$ de l'algorithme α , avec $y(\alpha(x)) \in Y$, soit maximum.

A noter également la notion de portfolio d'algorithmes [20] qui définit une exécution concurrente de plusieurs algorithmes résolvant un même problème.

d) Synthèse

Nous venons de présenter de manière concise les principales difficultés à résoudre lorsque l'on souhaite utiliser des méthodes statistiques pour prédire des séries temporelles.

Concernant les prétraitements des données, nous avons utilisé des méthodes classiques mais aussi développé des méthodes spécifiques prenant en compte la connaissance que nous avons du phénomène à modéliser. Le problème de la sélection des variables d'entrée du modèle a été abordé à l'aide de méthodes de type filtre. Nous avons ainsi utilisé des mesures entre variables basées sur la corrélation, l'auto-corrélation ou l'information mutuelle. Quant à la technique de l'hybridation, nous avons pu l'expérimenter en combinant plusieurs méthodes de prédiction suivant un critère de saisonnalité ou en suivant l'erreur de prédiction observée au pas de temps précédent. Les résultats de ces différentes applications sont l'objet de la prochaine section.

4. Applications au rayonnement solaire et aux pics de pollution

Dans cette section nous proposons de retracer les principaux résultats obtenus en termes d'expérimentations. Nous commencerons par le cas des données relatives au rayonnement solaire global (aussi appelé irradiation) et de production d'énergie photovoltaïque. Puis, nous présenterons les expériences menées pour estimer le rayonnement global incliné à partir du rayonnement global horizontal. Enfin, nous présenterons l'état de nos travaux concernant la prédiction de concentrations de polluants atmosphériques.

a) Prédiction de séries temporelles de rayonnement solaire global et de production d'énergie photovoltaïque à partir de réseaux de neurones artificiels

Ces travaux sur le rayonnement solaire global et la production d'énergie photovoltaïque correspondent au travail de thèse de M. Cyril VOYANT, thèse soutenue en 2011 [16]. Dans ce cadre nous avons axé notre étude sur le nord du bassin méditerranéen avec cinq sites d'étude (cf. figure ci-dessous) en testant des données en rapport direct ou indirect avec le rayonnement global.



Figure 7. Localisation des 5 villes étudiées.

Nous avons ainsi pu construire un outil de prédiction adaptable à différentes localités et à différents horizons. Les données que nous avons utilisées peuvent se classer en trois catégories :

- les variables endogènes : il s'agit des variables à prédire, éventuellement décalées dans le temps (rayonnement global ou puissance PV) ;

- les variables exogènes : il s'agit des variables différentes de la variable endogène (comme par exemple la température, la pression, etc.) ;
- les variables de prévisions : ce sont des variables exogènes non pas mesurées mais estimées à partir de modèles physiques.

Excepté les données de puissance PV (issues des façades inclinées de notre laboratoire et mesurées toutes les 5 min), l'ensemble de ces données ont été gracieusement mises à notre disposition par Météo-France dans le cadre de la résolution 40 sur l'échange international des données et produits météorologiques³.

Les caractéristiques principales des données de rayonnement global horizontal, des variables exogènes mesurées et de prévision dont nous disposons pour ces cinq villes sont indiquées sur le tableau suivant :

<i>Villes</i>	<i>Base</i>	<i>Période</i>	<i>Variables exogènes mesurées</i>	<i>Variables exogènes ALADIN</i>
Ajaccio	horaire	1971 à 1989 et 1998 à 2007	P,N,T,Ws,Wd, PKW,RP,RH,Su	P,N,RP,T
Bastia	horaire	1998 à 2007	idem	Idem
Marseille	horaire	1998 à 2007	idem	Idem
Montpellier	horaire	1998 à 2007	idem	Idem
Nice	horaire	1998 à 2007	idem	Idem

Tableau 1. Caractéristiques principales des données de rayonnement global horizontal.

Concernant les deux dernières colonnes du tableau, la notation suivante a été utilisée : P pour pression, N pour nébulosité, T pour température, Ws pour la vitesse du vent, Wd pour la direction du vent, PKW pour la vitesse de rafale, RP pour la précipitation, RH pour l'humidité et Su pour l'insolation.

A l'issu de ces manipulations, nous avons pu parfois constater des résultats totalement différents suivant l'horizon considéré. Durant ces travaux, notre approche a évolué : au départ, nous souhaitons optimiser les prédictors pour diminuer le plus possible l'erreur commise. Bien que ce type d'approche soit scientifiquement et techniquement intéressant, cette spécialisation des prédictors va à l'encontre de la généralisation des résultats. Dans une seconde étape, nous avons alors choisi de moins nous attarder à l'amélioration de la technique, mais à travailler sur une méthode qui serait utilisable par d'autres équipes. En effet, bien qu'il soit toujours intéressant de diminuer de quelques points de pourcentage le nRMSE d'une prédiction, il est fort probable que d'un point de vue énergétique les personnes qui

³ http://www.meteo.fr/meteo_france/resolution40/introduction.html

ont besoin de ces prédictions soient peu intéressées par le gain occasionné. Ce qui a été recherché avant tout est une méthode simple à mettre en place, rentable et exploitable sur plusieurs sites. En effet, la sélection des données et des paramètres des modèles choisis doit être parcimonieuse.

Le tableau ci-après propose une synthèse de ces manipulations en essayant de dégager le rapport efficacité/complexité de chaque prédicteur pour chaque horizon. Nous comparerons ensuite ces résultats avec ceux issus des différentes études bibliographiques.

<i>Horizon</i>	<i>Stationnarisation</i>	<i>Variables exogènes mesurées</i>	<i>Variables exogènes ALADIN</i>	<i>Prédicteur requis</i>	<i>Difficulté</i>	<i>Erreur (nRMSE)</i>
j+1	oui	Su,N,RH		PMC>ARMA>persistance	++	23,4%
h+1	oui		N,P,RP	Hybride PMC/ARMA > PMC>ARMA>persistance	+++	14,9%
h+24	oui			PMC>ARMA>persistance	+	27,3%
m+5	non			PMC> ARMA>persistance	+	20,2%

Tableau 2. Synthèse des résultats concernant le rayonnement solaire global et de production d'énergie photovoltaïque.

Dans le cas j+1, nous avons effectué des manipulations de prédiction de rayonnement global et d'énergie PV qui ont permis de mettre en place une méthodologie efficace de prédiction à base de PMC. De plus, nous avons remarqué qu'en l'absence de stationnarisation, il était équivalent d'utiliser un PMC ou une approche ARMA. Ces deux prédicteurs montrent des résultats globalement similaires. Pour ces deux approches, le fait de stationnariser apporte une valeur ajoutée à la qualité des prédictions. Nous obtenons un gain sur le nRMSE proche de 1 % par rapport au cas non stationnarisé. La dernière amélioration que nous avons obtenue concerne l'utilisation des variables exogènes. Néanmoins, ces dernières sont surtout efficaces durant les périodes à fortes occurrences nuageuses, sinon leur intérêt est minime. Le gain observé concernant le nRMSE est de ~1 % en mars et en avril et proche de 0 % en mai et en juin. Lors de l'étude j+1, nous avons aussi constaté que les PMC étaient capables d'utiliser des données d'apprentissage issues de zones géographiques différentes des zones de prédiction. Cette série de manipulation tend à confirmer le fait que les PMC, et plus particulièrement la méthodologie que nous avons utilisée, sont favorables à la prédiction du rayonnement global ou de la puissance PV.

Concernant le cas h+1, nous l'avons considéré comme l'horizon dont l'impact est le plus important pour un gestionnaire de réseaux électriques. C'est aussi celui qui

bénéficie le plus de l'apport des prédictions sophistiquées mises en œuvre. En effet, il n'y a que dans le cas horaire que nous estimons légitime d'utiliser un modèle complexe. Pour les autres horizons, la complexification des méthodologies induit des améliorations trop minimes pour les justifier. Le fait de stationnariser les séries temporelles permet de diminuer l'erreur de prédiction (nRMSE diminué de 1,7 %), et l'utilisation de variables exogènes (modèle de prévision ALADIN) a démontré une amélioration des résultats durant les tests (nRMSE diminué de 0,7 %). Cet horizon a surtout permis de comprendre l'intérêt de combiner les différents prédicteurs entre eux. Cette hybridation entre ARMA et PMC a donné de très bons résultats avec un nRMSE diminué de 3,5 %. La figure suivante propose une illustration générique de notre d'approche.

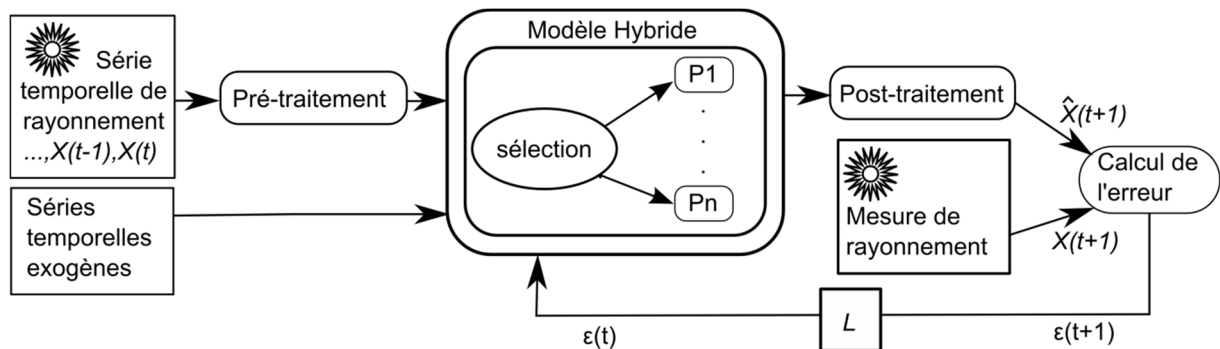


Figure 8. Principe de la méthodologie hybride.

Pour cet horizon, si l'on compare notre méthodologie globale (stationnarisation, variables exogènes et hybridation) avec une prédiction simple telle la persistance, on obtient une réduction de l'erreur de prédiction de plus de 11 %. La saison où le gain est le plus important est l'automne (nRMSE diminué de 16,1 %), suivi de l'hiver (nRMSE diminué de 15,8 %), de l'été (nRMSE diminué de 10,4 %) et du printemps (nRMSE diminué de 10 %).

Plus récemment [21], nous avons eu l'opportunité de faire évoluer cette approche en utilisant une approche bayésienne pour mixer différents modèles de prédiction. Si le PMC et ARMA se sont montrés équivalents (nRMSE près de 40,5% pour les deux), cette hybridation a permis un gain en % de nRMSE supérieur de 14 points par rapport à l'estimation de la persistance (nRMSE = 37% contre 51%). Toujours dans le cadre de cette étude, nous avons pu adopter une nouvelle méthode de sélection de variable plus robuste, en utilisant l'information mutuelle qui mesure la dépendance mutuelle entre deux variables. Plus l'information mutuelle est grande, plus les variables sont dépendantes. En effet, le coefficient de corrélation ou le coefficient de corrélation partielle, utilisé jusqu'à ce stade, indique uniquement la dépendance linéaire entre les variables. Or le PMC est dédié à l'estimation non linéaire et le lien entre les variables n'est pas forcément linéaire, ce qui distingue l'utilisation de l'information mutuelle. Les variables ainsi sélectionnées sont celles qui ont la plus grande l'information mutuelle avec la cible.

Enfin, pour conclure avec nos expériences sur cet horizon, nous avons encore plus récemment étudié [22] l'intérêt d'utiliser des données issues de satellites géostationnaires. En effet, lorsqu'un territoire est mal instrumenté, les données des satellites géostationnaires peuvent être utiles pour prédire le rayonnement solaire global. Nous avons obtenu un nRMSE de près de 16,5% pour les deux meilleurs prédicteurs (persistance et PMC).

Pour l'horizon h+24, les conclusions sont légèrement différentes. En effet, bien que la stationnarisation soit intéressante comme dans les deux cas précédents, nous n'avons pas pu trouver de réel intérêt à utiliser les variables exogènes. De plus, une combinaison de PMC proposée dans cette étude n'a pas conduit à de bons résultats. Il convient d'utiliser un PMC unique pourvu d'autant de sorties que de prédictions souhaitées (pour neuf heures, neuf neurones de sortie).

L'horizon m+5 remet également en cause ce qui a été observé pour les deux premiers horizons. En effet, cette fois-ci, c'est le mode de stationnarisation des séries temporelles qui n'est plus pertinent. Il est ainsi plus adéquat de manipuler les indices temporels en entrée du PMC. Ces indices permettent de diminuer le nRMSE de 0,7 %. Les résultats de prédiction de rayonnement global et de puissance PV sont améliorés par rapport à une approche classique, cependant le gain occasionné est minime (nRMSE diminué d'environ 1,5 %). En première approximation, il n'est certainement pas nécessaire d'utiliser un modèle abouti formé de PMC et d'indices temporels, une simple persistance peut suffire.

Dans tous les articles scientifiques relatifs à l'estimation du rayonnement global ou de l'énergie PV, on constate que les erreurs liées aux prédictions (mensuelles, journalières, horaires et minutes) diffèrent d'un site et d'un auteur à l'autre. Les méthodologies de prédictions sont, en règle générale, si différentes qu'elles sont difficiles à comparer. De plus, les estimations des erreurs sont hétérogènes : erreur de prédiction sur certains jours échantillonnés ou sur une large période, test effectué sur le cumul des prédictions, utilisation de paramètres d'erreur non usuels, différences de terminologie, etc. Pour s'affranchir de toutes ces spécificités qui entravent la comparaison, nous avons axé notre travail sur une méthodologie générale de l'estimation de l'erreur :

- test de prédiction sur une large période et non sur des journées « bien choisies » ;
- utilisation du RMSE pour pénaliser les gros écarts ;
- normalisation du RMSE pour pouvoir confronter les comparaisons sur plusieurs sites ;

- pas de cumul des prédictions (sauf études spécifiques) qui a pour effet de moyenner l'erreur et de la diminuer ;
- répartition des erreurs en fonction des saisons car la consommation électrique n'est pas la même tout au long de l'année ;
- tests réalisés sur plusieurs sites (dès que l'étude s'y prête), afin de s'affranchir des phénomènes climatologiques régionaux ;
- utilisation d'une prédiction test de référence et utilisation d'un prédicteur naïf afin de réellement statuer sur la qualité de notre méthodologie.

Cette méthodologie de l'estimation de l'erreur n'est pas celle utilisée dans la plupart des articles, il n'y a pas de réel consensus. A ce titre, il est difficile de comparer nos résultats avec ceux existants dans la littérature. Ce que nous présentons ici est une comparaison des différents prédicteurs élaborés et testés durant le travail doctoral de M. Cyril VOYANT afin d'en proposer une hiérarchisation.

Pour les horizons $j+1$ et $h+1$, nos résultats sont en partie conformes à la littérature [16]. En effet, les PMC sont adaptés et permettent d'effectuer des prédictions de rayonnement global avec une erreur acceptable et sont aussi applicables aux zones montagneuses. Concernant la hiérarchisation des modèles PMC et ARMA, nous nous démarquons des résultats bibliographiques traditionnels [16]. En Effet, sans étude approfondie sur la stationnarisation des séries temporelles utilisées, nous ne pensons pas qu'il soit aisé de différencier ARMA et PMC. Cependant nous sommes convaincus que l'utilisation des variables exogènes améliore les résultats des PMC. Comme dans la littérature, nous nous sommes aperçus que les approches pertinentes dans le cas de la prédiction du rayonnement l'étaient tout aussi dans le cas de la prédiction de la puissance PV. De plus, nous avons confirmé, dans le cadre de nos expérimentations, que les PMC pouvaient être utilisés quand la zone d'apprentissage et celle de test n'étaient pas identiques. Nous avons choisi la persistance comme prédicteur naïf. Bien qu'elle ne soit pas systématiquement utilisée dans la littérature, nous pensons que la persistance permet de juger correctement de la validité des techniques complexes.

Concernant l'hybridation des méthodes, nos conclusions restent identiques à ce que l'on peut trouver dans la littérature [16]. Enfin, concernant les horizons $h+24$ et $m+5$, bien qu'il existe encore trop peu d'études utilisant les PMC, nous pensons avoir montré que les PMC étaient adaptés à ces situations. De plus, notre approche avec l'utilisation des indices temporels semble efficace et mérite des approfondissements. En résumé, nos résultats sont relativement similaires à ceux que l'on trouve dans la littérature. Sous certains aspects ils sont complémentaires et, parfois, agrémentent la technique de prédiction d'outils novateurs.

b) Prédiction des concentrations de polluants atmosphériques en Corse et impacts sur la productivité des systèmes énergétiques solaires

L'objectif de ce travail s'inscrit dans le cadre des travaux de thèse de M. Wani TAMAS au sein de l'association agréée de surveillance de la qualité de l'air (AASQA) en Corse : « Qualitair Corse ». L'objectif est de réaliser un outil informatique de prédiction de concentration de polluants atmosphériques, capable d'effectuer des prévisions du jour au lendemain. Ces prévisions, effectuées par un modèle statistique sont localisées aux stations de mesure. A termes, ce modèle sera utilisé par Qualitair Corse afin d'anticiper la formation de pics de pollution et de permettre l'information des citoyens et des autorités ainsi que la prise de mesures préventives adaptées.

La prédiction de polluants est traditionnellement réalisée par des modèles de chimie-transport (CTM en anglais pour "Chemistry Transport Model") [23]. Ils sont proches dans le concept des NWP (cf. section VII.C.2.a), "Les modèles de prédiction numérique du temps"). Ce type de modèle divise l'atmosphère en plusieurs « boîtes », dans lesquels il calcule l'évolution de l'état atmosphérique, grâce à des équations décrivant les réactions chimiques et le transport des polluants, gaz ou particules. Ces modèles classiques de prédiction de la qualité de l'air fonctionnent mal en Corse pour plusieurs raisons : le relief de l'île et son impact sur la prévisibilité météorologique, le manque de données cadastrales d'émissions, l'absence de correction statistique des modèles existant mais aussi la position de la Corse en bordure des domaines de prédiction.

Ainsi, des modèles statistiques [24] reposant sur un apprentissage automatique des données disponibles en Corse et faciles à déployer, apparaissent comme une solution permettant d'améliorer la prédiction de la qualité de l'air, tout en demandant peu de ressources informatiques. Des principaux types de modèles statistiques, les réseaux de neurones artificiels et notamment les PMC, ressortent particulièrement pour leur non-linéarité, leur apportant la capacité de modéliser des liens complexes entre les données.

Dans ces travaux les données utilisées (voir figure et tableau ci-après) sont des séries temporelles horaires de variables provenant de plusieurs sources. Qualitair Corse fournit les mesures de concentrations des polluants (l'ozone (O_3), les particules fines de diamètre aérodynamique inférieur à $10\mu m$ et $2.5\mu m$ (PM10 et PM2.5), les oxydes d'azotes (NO et NO_2), le dioxyde de soufre (SO_2)). Les variables météorologiques (direction et vitesse du vent, précipitations, rayonnement global, humidité relative, température, pression) proviennent soit des stations de Qualitair Corse soit de stations de Météo-France.

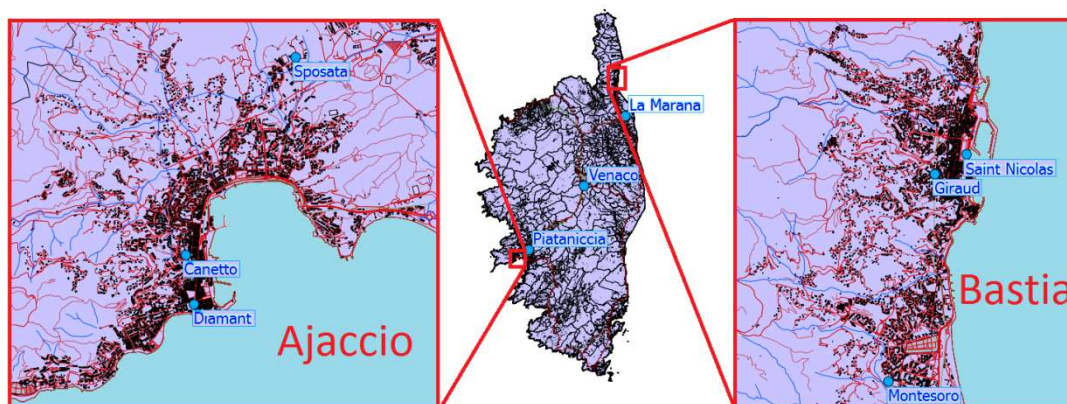


Figure 9. Situation géographique des stations de surveillance en Corse (fournie par l'IGN).

Ville	Stations	Catégorie	Hauteur niveau moyen de la mer	Polluants mesurés
Bastia	Giraud	Urbain	60 m	NO ₂ O ₃ PM ₁₀ SO ₂
	Montesoro	périurbain	47 m	NO ₂ O ₃ PM _{2.5}
	St Nicolas	Trafic	5 m	NO ₂ PM ₁₀
	La Marana	Industriel	15 m	NO ₂ O ₃ PM ₁₀ SO ₂
Venaco	Venaco	Rural	653 m	NO ₂ O ₃ PM ₁₀ PM _{2.5}
Ajaccio	Canetto	Urbain	39 m	NO ₂ O ₃ PM ₁₀ SO ₂
	Sposata	périurbain	60 m	NO ₂ O ₃
	Piatanaccia	Industriel	30 m	NO ₂ O ₃ PM ₁₀ SO ₂
	Diamant	Trafic	12 m	NO ₂ PM ₁₀

Tableau 3. Descriptif des stations de mesure en Corse.

Nous avons également utilisé des sorties de modèles de prévision météorologique, principalement le modèle AROME de Météo-France, qui fournissent les mêmes variables que celles disponibles en station sur plusieurs niveaux. On utilise également d'autres variables comme la nébulosité totale, la hauteur de la couche limite ou des vitesses de vent verticales que fournit AROME.

Certaines expériences nous ont amenées à utiliser des données fournies par Air PACA, l'AASQA de la région PACA, d'où peuvent-nous parvenir des masses d'air en situation de mistral. Air Paca gère également une plate-forme de modélisation nommée AIRES, qui fournit des prévisions de concentrations en polluants sur la Corse. Nous avons pu récupérer les sorties de cette plate-forme, et les utiliser, soit pour alimenter notre modèle, soit à des fins de comparaison.

La prédiction s'est axée principalement autour de deux polluants : l'O₃ puis les PM₁₀. Les premiers modèles de prévision de l'O₃ permettent la prévision 24 heures à l'avance. Ils ont été construits en suivant la méthodologie développée pour les séries temporelles de rayonnement. Nous avons ainsi atteint une précision intéressante, comparable avec celle qu'on peut trouver dans la littérature pour ce type de prévision statistique de l'ozone [25][26]. Il outrepassa de loin la précision d'AIRES, le CTM développé par Air PACA, basé sur le modèle méso-échelle CHIMERE couvrant le domaine de la Corse.

Les deux figures suivantes illustrent le résultat des prédictions sur quelques jours, à Bastia et Ajaccio.

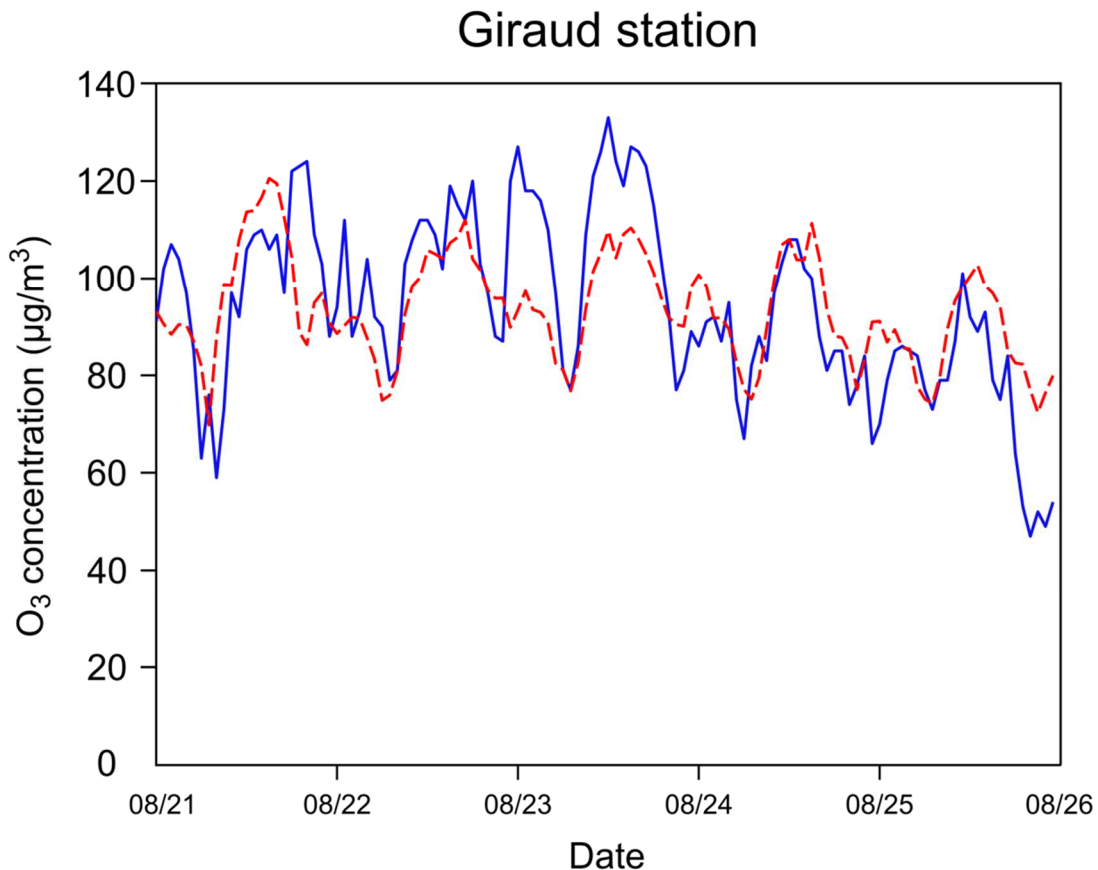


Figure 10. Concentration d'ozone observée (en bleue) et prévue 24h à l'avance (en rouge) en août 2013 à la station Giraud (Bastia).

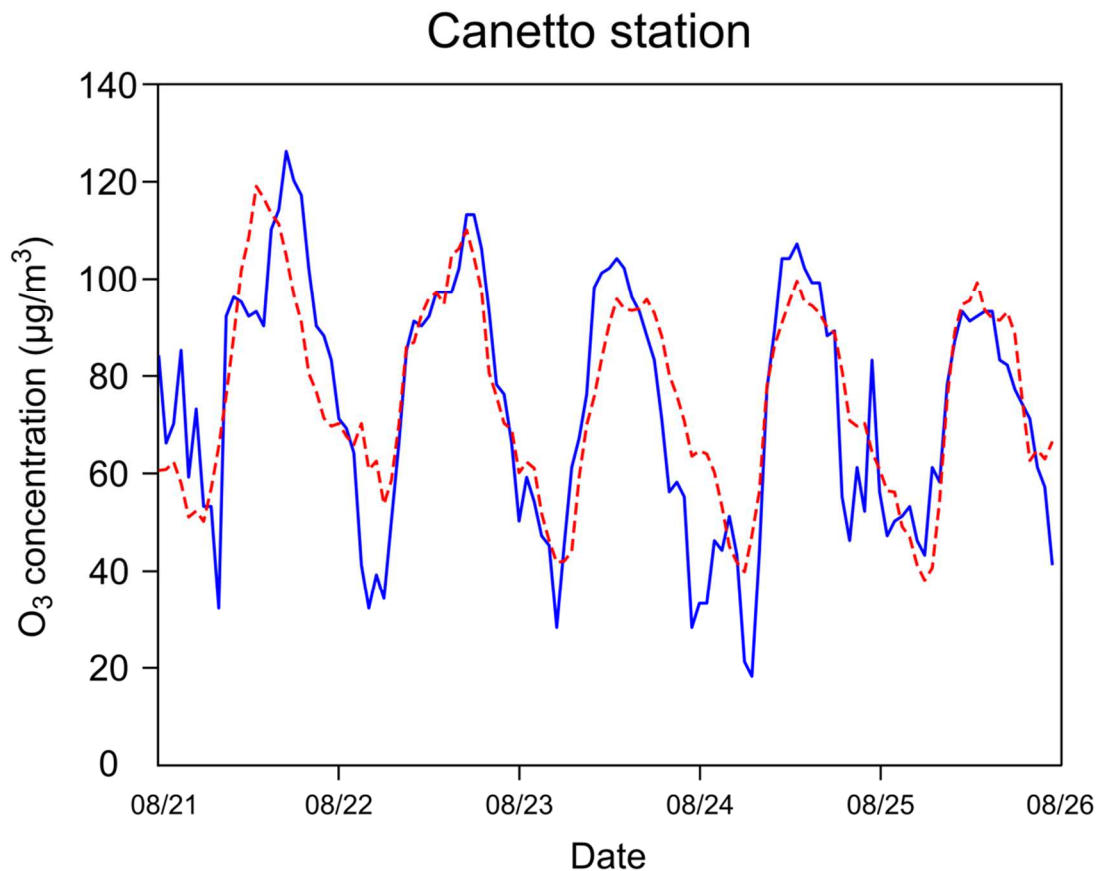


Figure 11. Concentration d'ozone observée (en bleue) et prévue 24h à l'avance (en rouge) en août 2013 à la station Canetto (Ajaccio).

Les scores sont logiquement meilleurs que ceux du CTM AIRES. La précision des prévisions de PM10 est moins bonne que celle d'ozone, en raison de la complexité du polluant. On note que la RMSE est plus petite dans le cas de PM10 à cause des valeurs de concentration du polluant qui sont plus basses que pour l'ozone. La nRMSE et l'IA indiquent que les prévisions d'O₃ sont meilleures que celles des PM10. En effet, les particules fines ont une dynamique complexe et sont plus difficiles à prévoir, que cela soit avec des modèles statistiques ou des CTM. Ces résultats sont satisfaisants quand on les compare à ceux obtenus récemment avec des réseaux de neurones artificiels dans des villes comme Phoenix en Arizona [27].

	RMSE (µg.m ³)	nRMSE (%)	MAE(µg.m ³)	IA
PM10 avec PMC	7.16	37.99	5.31	0.68
PM10 avec Aires	13.37	70.93	10.70	0.55
O ₃ avec PMC	17.85	20.38	13.33	0.71
O ₃ avec Aires	37.51	42.82	34.12	0.42

Tableau 4. Performances de la prévision à J + 1 des modèles basés sur les PMC et du modèle AIRES, pour l'O₃ et les PM10.

L'objectif de l'apprentissage du PMC par l'algorithme de Levenberg-Marquardt étant de minimiser la MSE, le modèle entraîné a une bonne précision moyenne, mais est moins précis lors d'événements rares avec des concentrations en polluants importantes.

Etant donnée la nécessité d'anticipation des pics pour la santé publique, la sous-estimation des pics reste un point faible de ces prédictions. La suite des études se focalisera sur ce point. En effet, afin d'obtenir de meilleurs prévisions des pics de pollution, nous travaillons actuellement sur des modèles de classification, qui permettent de séparer les données en différents groupes, chacun représentatif d'un régime atmosphérique particulier.

Cette approche est à rapprocher de la technique d'hybridation de modèle présentée dans les travaux précédents sur le rayonnement. Après cette classification, il sera possible d'entraîner un PMC pour chaque régime. On aura ainsi plusieurs modèles, et pour chaque prévision on utilisera celui qui correspondra au régime considéré [28][29]. L'entraînement d'un PMC uniquement sur les données correspondant aux régimes propices aux pics de pollution peut améliorer la précision lors des pics.

La figure suivante illustre ce principe qui permet ainsi d'avoir des modèles spécialisés sur les régimes qui favorisent les pics de pollution, avec un gain de précision attendu.

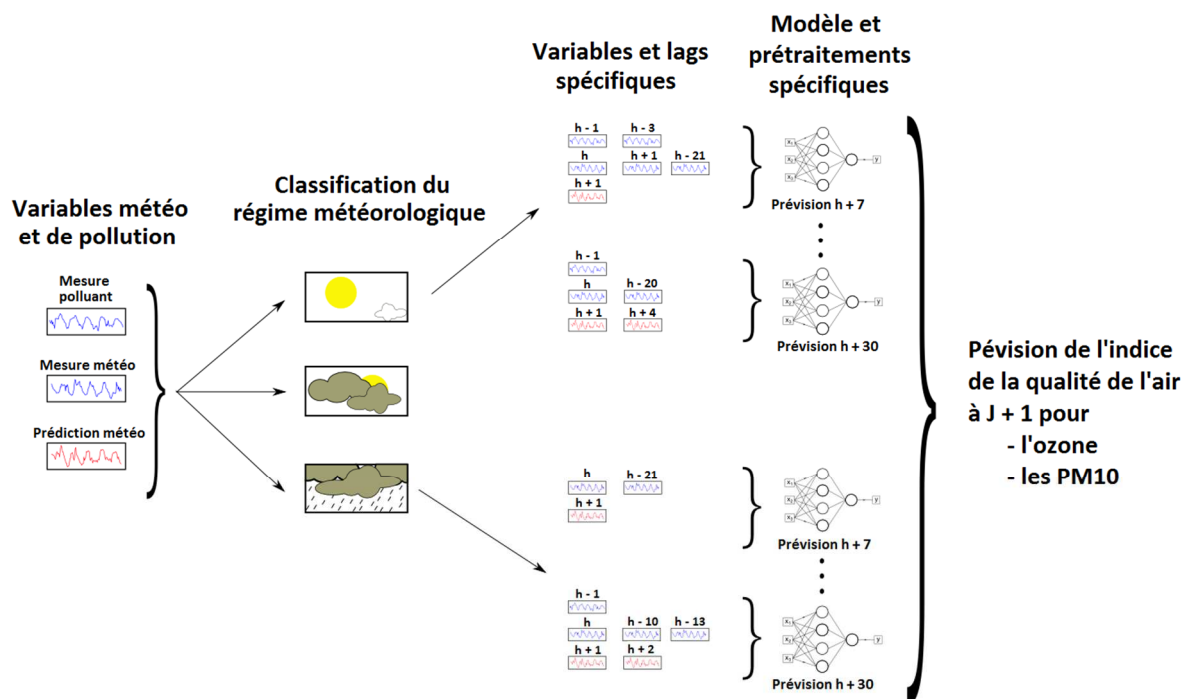


Figure 12. Principe des modèles prévisionnels avec classification.

La classification des régimes météorologiques peut elle-même se faire selon deux méthodes. La première consiste à utiliser un modèle statistique de partitionnement (clustering), qui va diviser les jeux de données en différents groupes, chacun constituant une classe. L'autre méthode se base sur les connaissances aérologiques de l'utilisateur, qui crée les différentes classes pour correspondre aux régimes météorologiques qu'il a identifié.

Dans tous les cas il est important de limiter le nombre de classes, pour qu'il subsiste un nombre de données de chaque classe suffisant pour assurer l'apprentissage de chaque modèle dans de bonnes conditions.

Chaque modèle spécifique à une classe peut être monté différemment des autres. On peut utiliser des variables différentes, des prétraitements différents et une architecture différente afin de mieux adapter chaque modèle aux interactions spécifiques entre les variables pour son régime météorologique.

Le second volet de ce travail doctoral concerne l'étude des impacts de la pollution sur la productivité des systèmes énergétiques solaires. Non encore initiés à ce stade, ces travaux constituent de facto une perspective.

c) Utilisation des réseaux de neurones artificiels pour l'estimation du potentiel solaire

Initiée par deux stages de Master en 2011, cette étude se poursuit actuellement avec le travail doctoral de Mlle Kahina DAHMANI, démarré en 2012.

Dans cette étude, nous proposons d'utiliser les réseaux de neurones artificiels, non plus pour prédire, mais pour estimer le rayonnement solaire incliné à partir du rayonnement solaire global horizontal et d'autres variables météorologiques disponibles tels que la température ambiante, la vitesse du vent, la nébulosité, etc.

Comme le souligne Behr [30], trois raisons principales font qu'il est très difficile de développer un modèle simple qui permette de convertir l'irradiation solaire reçue sur un plan horizontal en l'irradiation incidente sur un plan incliné :

- la radiation solaire incidente sur un plan incliné inclut la radiation réfléchie par l'environnement ;
- lorsque la surface est inclinée, elle ne « voit » qu'une partie du ciel et donc la partie diffuse du rayonnement solaire incidente sur la surface est liée à l'angle solide qui en découle ; cette irradiation solaire en provenance du ciel ne dépend pas que de l'angle d'inclinaison ou de l'orientation du capteur, de la hauteur du soleil dans le ciel et de son azimut mais également de l'état du ciel qui est rarement uniforme et donc induit des effets d'anisotropie difficilement quantifiables et complexes. Il y a deux principaux problèmes dans cette anisotropie du ciel : la brillance circumsolaire due à la diffusion de la radiation solaire par les aérosols et concentrée dans la zone du ciel autour du soleil et la brillance de l'horizon concentrée près de l'horizon et plus prononcée par ciel clair ;
- enfin, ces données sont relativement rares.

Quand la radiation solaire pénètre dans l'atmosphère terrestre, une part de l'énergie incidente est renvoyée vers l'espace, une partie est diffusée et absorbée (radiation diffuse) et une partie atteint directement le sol (radiation directe). La radiation arrivant sur une surface inclinée possède ainsi une composante directe (nulle par ciel nuageux) et deux composantes diffuses du ciel et du sol (voir figure ci-après).

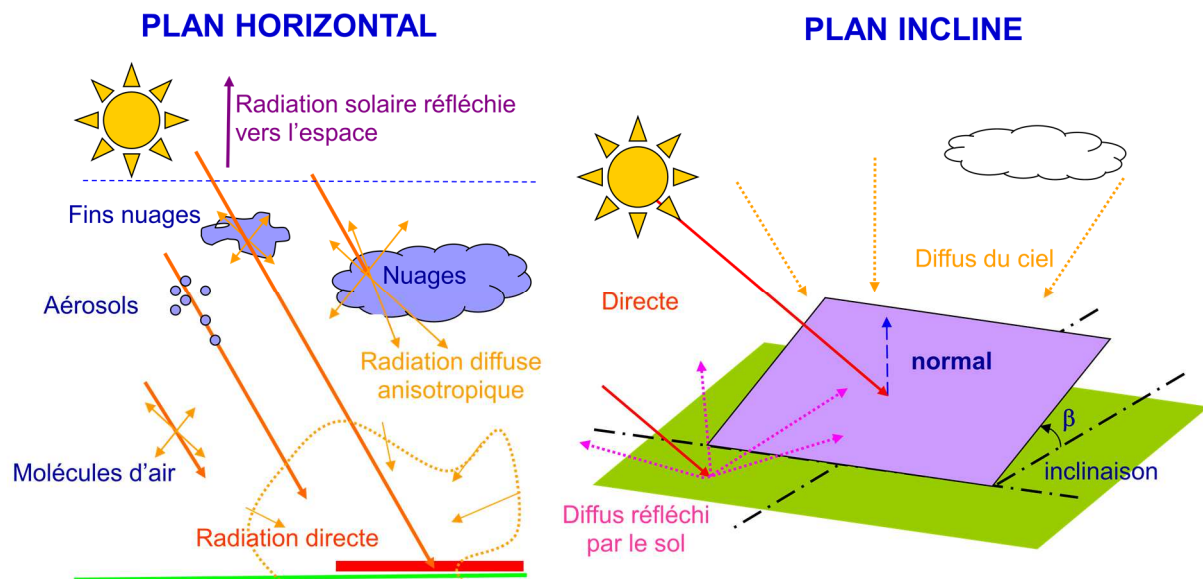


Figure 13. Radiation incidente sur une surface horizontale et inclinée.

Les relations entre la radiation horizontale et les autres composantes ne sont pas linéaires et dépendent :

- de la position du soleil ;
- de l'état du ciel: densité et position des nuages, concentration en vapeur d'eau, etc.

De plus, une surface inclinée ne voit qu'une partie du ciel et celui-ci n'étant pas uniforme, la radiation diffuse est anisotrope et il n'existe donc pas de relation biunivoque entre l'irradiation diffuse inclinée et horizontale. Les réseaux de neurones artificiels nous ont donc paru particulièrement bien adaptés pour exploiter des relations non linéaires entre ces composantes.

Les données utilisées proviennent de deux sites de mesures. Le premier est le laboratoire CNRS UMR 6134 (Vignola) de l'Université de Corse dans le golfe d'Ajaccio (latitude : $41^{\circ}55'$ N ; longitude : $8^{\circ}55'$ E), situé à environ 200 m de la mer et à 70 m d'altitude. Ce site est caractérisé par un climat méditerranéen insulaire. Le second site est le Centre de Développement des Energies Renouvelables (CDER) disposant d'une station météorologique à Bouzareah, près d'Alger (latitude : $36^{\circ}8'$ N ; longitude : $3^{\circ}17'$ E) à une altitude de 347 m. Le site est caractérisé par un climat méditerranéen avec des étés chauds et secs et des hivers humides et frais.

Les caractéristiques principales des données disponibles sont indiquées sur le tableau suivant :

Villes	Base	Période	Inclinaison
Vignola/Ajaccio	horaire, 10 min	depuis 2007	45° et 60°
Bouzareah/Alger	5 min	avril 2011 à mars 2013	36,8°

Tableau 5. Caractéristiques principales des données disponibles à Vignola et Bouzareah.

En plus de la variable mesurée I , l'irradiation solaire globale horizontale (Wh/m^2), nous avons souhaité disposer pour chacun des 2 sites, de quatre paramètres calculés :

- I_0 , l'irradiation solaire horizontale hors-atmosphère (Wh/m^2) ;
- δ , la déclinaison du soleil (rad) ;
- θ_z , l'angle zénithal (rad) ;
- ψ , l'angle azimut (rad).

Trois séries d'expérimentation ont pu être menées à ce jour avec ces quatre paramètres ajoutés en entrée d'un réseau de neurones artificiels de type PMC. Ces résultats peuvent être résumés comme suit :

- Base 10 min à Vignola/Ajaccio : Le PMC a été optimisé en utilisant cinq années de données solaires avec un nRMSE d'environ 9% pour la configuration optimale. Ce résultat est similaire ou légèrement inférieur aux erreurs constatées dans la littérature pour l'estimation de données horaires.
- Base horaire à Vignola/Ajaccio : Le PMC a été optimisé sur le même historique que précédemment en obtenant un nRMSE d'environ 6% pour la configuration optimale. Ce résultat est inférieur aux erreurs constatées dans la littérature.
- Base 5 min à Bouzareah/Alger : Le PMC a été optimisé en utilisant deux années de données solaires avec un nRMSE d'environ 8% pour la configuration optimale. Ce résultat confirme les bons résultats obtenus pour Vignola/Ajaccio sur une base de 10 min.



Université de Corse - Pasquale Paoli
Université Galatasaray
-
UMR CNRS 6134
Sciences Pour l'Environnement
Ministère des Affaires Etrangères



Très prochainement nous aurons l'occasion de tester l'intérêt d'autres variables telles que les variables météorologiques comme la température ambiante, la vitesse du vent ou encore la nébulosité.

5. Synthèse

Cette section conclue le premier axe de mes travaux de recherche développé au sein de l'équipe EnR de l'UMR 6134 ; et qui a concerné la modélisation et la prédiction de phénomènes spatio-temporels.

Nous nous sommes plus particulièrement intéressés à l'utilisation des réseaux de neurones artificiels pour prédire des séries temporelles d'ensoleillement et de concentration en polluants atmosphériques. Nous avons étudié tout d'abord l'impact de prétraitements statistiques. Puis, au vu du nombre important de données dont nous disposions, nous avons étudié différentes méthodes de sélection de caractéristiques (ou de variables) avec l'objectif d'obtenir le modèle le plus performant mais aussi le plus simple possible (principe de parcimonie). Enfin, nous avons étudié comment hybrider différentes méthodes de prédiction.

Plus récemment nous avons démarré des travaux concernant l'estimation de potentiel solaire également à l'aide réseaux de neurones artificiels. Il s'agit cette fois non pas de prédire mais d'estimer le rayonnement solaire incliné à partir du rayonnement solaire global horizontal et d'autres variables météorologiques disponibles.

Les principaux résultats obtenus en termes d'expérimentations ont pu à chaque fois être présentés et commentés. Nous avons commencé par le cas des données relatives au rayonnement solaire global (aussi appelé irradiation) et de production d'énergie photovoltaïque. Puis nous avons présenté l'état de nos travaux concernant la prédiction de concentrations de polluants atmosphériques. Enfin, nous avons présenté les expériences menées pour estimer le rayonnement incliné à partir d'un rayonnement global horizontal et de paramètres calculés.

Le prochain paragraphe est consacré au second axe de recherche que j'ai pu développer au sein de l'équipe EnR. Celui-ci propose d'étudier l'apport des langages informatiques de représentation des connaissances pour représenter la connaissance produite tout au long des processus d'aide à la décision. Des projets tels que les centrales photovoltaïques ou encore les fermes éoliennes constituent des exemples significatifs de processus de décision extrêmement complexes, difficiles et conflictuels.

D. Aide à la décision

La recherche d'un processus d'aide à la décision efficient peut s'avérer particulièrement complexe et difficile lors du montage et la réalisation de projets de grande dimension définis à différentes échelles et sur différentes localisations du territoire. Des projets tels que les centrales photovoltaïques ou encore les fermes éoliennes constituent des exemples significatifs. Trop souvent, la question du conflit et celle de l'acceptation sociale par la population de ce type de projets, est sous-estimée.

Dans ce contexte, nous pensons qu'un bon usage des outils et technologies informatiques, mais aussi la construction de nouveaux logiciels doivent permettre de répondre à des impératifs de gouvernance plus transparente et plus participative particulièrement ancrée à l'échelon local. On retrouve dans la littérature souvent le terme de « e-gouvernance » pour identifier le domaine d'appartenance de ces outils.

La mise en œuvre de ce projet est exposé ci-après en cinq points : (i) concepts de base de l'Aide MultiCritère à la Décision (AMCD) ; (ii) état de l'art concis et constats ; (iii) concepts de base du web sémantique, (iv) représentation de la connaissance par des langages dédiés ; (v) synthèse et perspectives projetées.

1. Concepts de base de l'Aide MultiCritère à la Décision

L'Aide MultiCritère à la Décision (AMCD) est une branche de la recherche opérationnelle considérée comme un outil d'aide à la décision. En effet, l'objectif de la recherche opérationnelle est de déterminer des méthodes et techniques scientifiques (mathématiques et informatiques) en recherchant la meilleure façon d'effectuer des choix en vue d'aboutir au résultat visé ou au meilleur résultat possible.

Contrairement au calcul économique des choix et à la recherche opérationnelle classique, les modélisations de type AMCD permettent de proposer des recommandations basées sur la recherche de solutions de compromis, satisfaisantes plutôt qu'optimales [31]. L'approche multicritère apporte ainsi des solutions au problème de l'évaluation selon différents points de vue souvent contradictoires [32]. Elle consiste en la définition et l'étude approfondie de critères reflétant les différents points de vue entrant en considération dans le processus de décision. On crée ainsi une matrice des performances qui synthétise les données d'entrée de l'analyse multicritère.

Critère					
		j_1	j_2	...	j_n
Alternative	a_1	$g_1(a_1)$	$g_2(a_1)$...	$g_n(a_1)$
	a_2	$g_1(a_1)$	$g_2(a_2)$...	$g_n(a_1)$

	a_n	$g_1(a_n)$	$g_2(a_n)$...	$g_n(a_n)$

Tableau 6. Structure de la matrice de performance.

L'évaluation de la performance des actions a_i à comparer pour les critères j est classiquement notée $g_j(a_i)$. De façon générale, le processus d'évaluation des performances des actions au regard de l'ensemble des critères peut se faire selon plusieurs modes, laissant plus ou moins au décideur la possibilité d'exprimer ses préférences :

- la notation : attribution d'une note sur une échelle de valeurs ;
- l'évaluation numérique : quantification numérique d'un impact ou d'une valeur.

Chaque critère j peut être accompagné de seuils (indifférence, préférence ou veto par exemple) permettant de définir des zones de préférence entre deux actions.

Il est également demandé aux différents évaluateurs, appelés acteurs, de pondérer les critères, permettant ainsi de donner un poids plus important à un ou plusieurs critères parmi le jeu initial de critères.

La clé de voute des approches multicritères réside dans l'opération mathématique, appelée agrégation des performances, consistant à faire la synthèse des informations regroupées dans la matrice des performances. Plusieurs types d'approches opérationnelles existent dans la littérature et dépendent du contexte d'utilisation et de la démarche choisie. On distingue ainsi les démarches descriptives, lorsqu'on ne souhaite pas perturber le processus de décision, des démarches constructives qui font l'hypothèse dès le départ de conflits forts et inévitables.

Le résultat de cette agrégation est représenté en général sous forme d'un graphe représentant les relations entre les actions à comparer, appelées relations de surclassement.

L'intérêt de ces méthodes a largement été démontré au sein de notre équipe par des travaux sur la localisation de parcs éoliens en région Corse [31] ou encore sur la sélection de projets photovoltaïques sur des terres agricoles [33].

2. Etat de l'art et constats

Malgré une approche qui se veut transparente, de type "boîte blanche", les méthodes multicritères présentent paradoxalement l'inconvénient d'être peu lisibles aux yeux de l'utilisateur non spécialiste. En effet, leur utilisation plus forte dans le milieu non académique spécialiste est freinée par des difficultés de compréhension d'un vocabulaire, de techniques et de variantes très riches et complexes. Ces difficultés peuvent conduire à un manque d'intelligibilité des résultats pour les décideurs, voir à une défiance et une méfiance vis à vis de ce type de méthodes. Suite à un premier état de l'art sur cette problématique publié lors des 72^{èmes} journées internationales de l'AMCD [34], nous avons pu établir entre autres les constats suivants :

1. l'utilisation plus forte de l'AMCD dans le milieu non académique spécialiste est freinée par des difficultés de compréhension d'un vocabulaire très riche et complexe. Ces difficultés peuvent conduire à un manque d'intelligibilité des résultats pour les décideurs, voir à une défiance et une méfiance vis à vis de la méthode ;
2. la possibilité de réutilisation de travaux antérieurs sur des processus de décision similaires reste faible. Des efforts de standardisation des données manipulées et des méthodes de calcul ont été réalisés mais quasiment aucun sur des vocabulaires de représentation de la connaissance produite ;
3. le besoin de plus en plus fort de connecter l'AMCD à des outils informatiques tel que les Systèmes d'information géographique (SIG), demande une certaine clarification du sens des données manipulées afin d'offrir une interopérabilité optimale.

Conforté dans notre désir d'étudier le bénéfice des technologies issues du web sémantique permettant de représenter la connaissance, nous nous sommes orientés vers les langages de représentation des connaissances développés par le W3C (organisme de standardisation du web).

3. Concepts de base du web sémantique

Inventé par Tim Berners-Lee (également père du web classique), le principe du web sémantique a été présenté pour la première fois en 1994 à la conférence "International World Wide Web Conference". Son objectif est de rendre compréhensible le web actuel par les machines, qu'elles soient de type logiciel ou matériel, afin que celles-ci puissent inférer à la place des humains.

Tim Berners-Lee propose ainsi, non pas de remplacer le web actuel, mais d'ajouter une couche expressive de connaissance. La figure⁴ suivante, illustre ce concept.

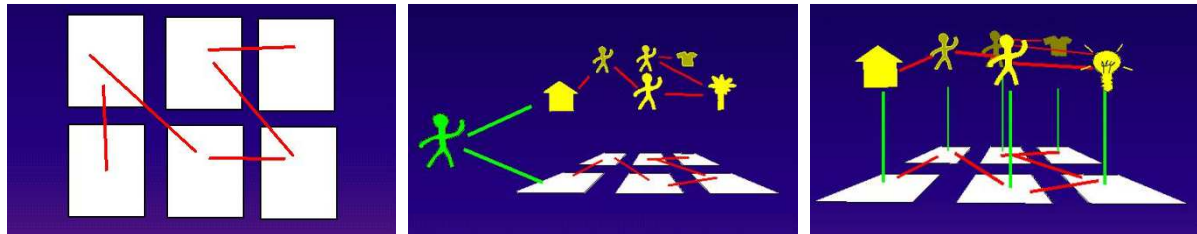


Figure 14. Différence entre le web actuel et le web sémantique.

Sur la partie gauche de la figure est illustré le web classique. Il s'agit d'hyperliens entre documents ou pages écrits en HTML, le tout formant un immense graphe. La figure du milieu illustre la façon dont les humains naviguent à travers ces pages. Grâce à leur connaissance, les humains interprètent ce qu'ils lisent, voient ou entendent et suivent les liens en fonction de leur désir ou besoin. L'idée de Tim Berners-Lee est alors illustrée sur la figure de droite. Il propose de connecter la connaissance des humains aux pages web.

Techniquement cela revient à décrire cette connaissance à l'aide d'un système de métadonnées formelles, appelés également vocabulaires ou ontologies, puis à la connecter, voir l'insérer dans les pages web.

Depuis 1994, cette vision a bien avancée puisque des standards reconnus par le W3C et s'ajoutant à ceux du web classique, forment l'ossature du web sémantique :

- RDF : modèle conceptuel permettant de décrire toute donnée ;
- RDF Schema : langage permettant de créer des vocabulaires, ensembles de termes utilisés pour décrire des choses ;
- OWL : langage permettant de créer des ontologies, vocabulaires plus complexes servant de support aux traitements logiques (inférences, classification automatique ...) ;
- SPARQL : langage de requêtes pour obtenir des informations à partir de graphes RDF ;
- RIF : langages de règles pour représenter une connaissance.

⁴ Issue de <http://www.w3.org/Talks/WWW94Tim/>

4. Représentation de la connaissance avec les langages RDF-RDFS, OWL et SKOS

Les langages permettant de construire des ontologies comme RDF-RDFS et OWL ajoutent une couche expressive très puissante au langage RDF⁵, et fournissent des outils permettant de définir des structures conceptuelles très complexes. Cependant, cette modélisation précise et basée sur des classes est exigeante en termes de compétences et d'effort. Dans de nombreux cas, il s'est même avéré que ce type de modélisation peut être superflu ou inadapté aux exigences. Ainsi parmi les langages sémantiques du W3C, SKOS (Simple Knowledge Organization System) a tout d'abord retenu notre attention pour son adéquation avec nos besoins. Son niveau d'abstraction intermédiaire entre RDF-RDFS et OWL facilite l'accès aux taxonomies et ontologies légères. De plus les recommandations SKOS mettent en avant le traitement multilingue de concepts organisés. Enfin, un autre avantage de SKOS est qu'il peut être utilisé en combinaison avec les langages d'ontologie Web tel que OWL.

Ainsi orienté vers le langage SKOS, nous avons formalisé un glossaire français et anglais édité au sein du Groupe de Travail Européen de l'AMCD par B. Roy, professeur émérite à l'Université Paris Dauphine. Une fois validé par la communauté et diffusé au travers des sites web du W3C, ce vocabulaire sera mis à disposition librement pour être réutilisé au sein de toute application informatique traitant de l'AMCD mais également au sein de document. Un des premiers effets sera de favoriser la recherche d'information sur le sujet au sein des moteurs de recherche web qui intègrent de plus en plus les technologies du web sémantique. Un autre effet sera d'améliorer la productivité des équipes de développeurs du domaine de l'AMCD grâce à une meilleure compréhension des données et termes manipulés relatifs à celui-ci.

5. Synthèse et perspectives

Dans un premier temps, il nous a donc été possible de clarifier la problématique proposée par un premier état de l'art. Aussi le langage SKOS a été identifié pour la construction d'un vocabulaire de l'AMCD. Nous avons proposé de poursuivre ces efforts afin de valider et diffuser ce vocabulaire ainsi que de démarrer un travail doctoral.

Concernant notre vocabulaire décrit en SKOS, nous en avons fait la présentation et publié une version beta lors des 73^{èmes} journées de l'AMCD⁶ couplées aux 8^{èmes} journées de travail du projet « Decision Deck » en avril 2011. Les possibilités d'intégration et apport de ce vocabulaire au sein de la plateforme logiciel « Decision

⁵ RDF (Resource Description Framework) est un modèle de graphe destiné à décrire de façon formelle les ressources Web et leurs métadonnées, de façon à permettre le traitement automatique de telles descriptions.

⁶ <http://mcda.univ-corse.fr>

Deck » ont également été évoquées. Ces journées étaient dédiées aux approches spatiales de l'évaluation multicritère et donc à l'apport des SIG à l'AMCD. Il est à noter que notre université a eu l'honneur d'organiser ces journées et de recevoir l'ensemble des membres de cette communauté, en intégrant le comité scientifique et d'organisation de ces deux manifestations et en animant leur site web.

Suite à la présentation d'une communication sur un premier état de l'art de la problématique ici exposée [34], un sujet de doctorat en collaboration avec notre collègue Pascal OBERTI et José R. FIGUEIRA, Professeur à l'Université Technique de Lisbonne et collaborateur au LORIA (laboratoire lorrain de recherche en informatique et ses applications, Unité Mixte de Recherche 7503), a pu être défini. Ce dernier a largement étudié le niveau d'intégration et la complexité des SIG au sein de l'AMCD. Le sujet est intitulé : « Evaluation multicritère et E-Gouvernance territoriale de projets d'énergies renouvelables: construction des connaissances d'acteurs au sein d'un système d'information géographique et sémantique ».

Bien que ciblé sur les projets d'énergies renouvelables, le travail doctoral visera une formalisation générique d'un processus d'aide à la décision. Nous avons ainsi identifié le besoin de définir l'ensemble des concepts de l'AMCD par le biais d'une ontologie spécifique ajoutant une couche de connaissance aux bases de données et ontologies existantes. Les principales tâches de ce travail doctoral ont été listées comme suit :

1. état de l'art et étude des ontologies existantes en relation avec l'AMCD ;
2. analyse ontologique d'un ou plusieurs processus décisionnels de type énergétique ;
3. état de l'art et étude des architectures technologiques pour construire et maintenir un système d'information sémantique et géographique ;
4. proposition d'une architecture technique et développement logiciel.

A l'issue de ce travail doctoral, nous pensons être en mesure de mettre à disposition un outil logiciel permettant : la formalisation de processus de prise de décision, l'amélioration de la qualité de la gouvernance, la facilitation de l'accès à l'information pour la prise de décision des acteurs, l'amélioration de la connaissance des données disponibles pour les décideurs, la facilitation de la recherche automatique d'informations.

6. Conclusion

Nous avons souhaité réunir au sein d'une approche innovante et pluridisciplinaire informaticiens, physiciens et économistes autour de plusieurs compétences : recherche opérationnelle et Aide MultiCritère à la Décision (AMCD), représentation des connaissances et ontologies développées dans le cadre du web sémantique, Systèmes d'Information Géographiques (SIG), Energies nouvelles et Renouvelables (EnR).

Ce projet devrait permettre de positionner l'Université de Corse au sein des très récentes démarches TIC, dites « GreenIT », visant à concevoir ou à employer celles-ci pour réduire les effets négatifs de l'activité humaine sur l'environnement. Est visée plus particulièrement la branche « Sustainable IT » dédiée à l'amélioration des conditions de vie des citoyens et facilitant leur participation aux enjeux et à la gouvernance de la société par l'usage des TIC. Enfin il permettra à l'université de Corse de s'ouvrir aux thématiques de recherche liées au web sémantique et aux travaux du W3C.

Plus prospective, moins volumineuse en terme de production et de collaborations scientifiques par rapport à la thématique précédente, cette thématique ne sera pas abordée pas dans la conclusion générale de ce mémoire.

E. Test de logiciel

Dans cette partie qui conclue la synthèse de mes travaux de recherche, il s'agit de retracer des travaux plus anciens que les deux précédemment traités, mis en veille depuis mon intégration dans l'équipe EnR en 2005. Ces travaux concernent l'application de techniques de test de logiciel pour le test de circuits électroniques décrits en VHDL qui est un langage de programmation spécialement conçu pour modéliser et simuler des circuits électroniques.

1. Problématique

Typiquement, les concepteurs de circuits commencent par l'écriture, à partir de spécifications textuelles (sous la forme d'un cahier des charges), d'une description au niveau algorithmique ou RTL de la fonctionnalité du système en utilisant un langage de description de matériel (HDL en anglais pour « Hardware Description Language »). Ensuite, les outils de synthèse sont utilisés pour transformer la description HDL de niveau RTL en une description au niveau porte logique, puis au niveau transistor, et produisent finalement le masque au niveau « layout » qui est employé pour fabriquer le circuit. Des outils de CAO (Conception Assisté par Ordinateur) sont utilisés intensivement pour exécuter les transformations entre les différents niveaux d'abstraction. Ils aident également à optimiser la description afin de prendre en compte les contraintes de vitesse, de consommation de puissance, et de surface.

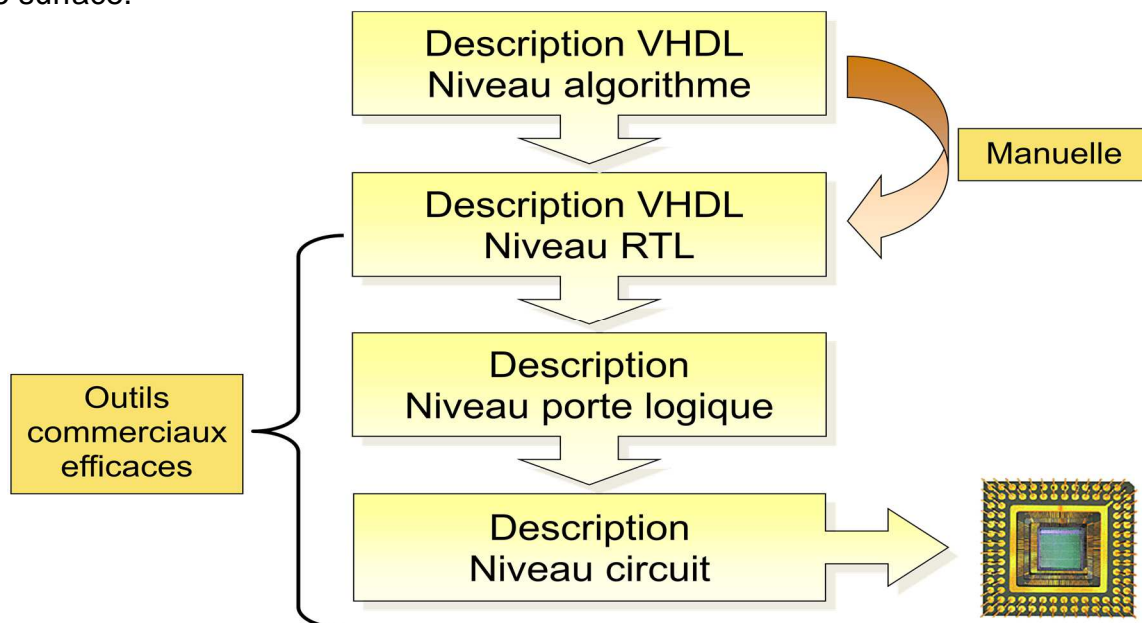


Figure 15. Processus de conception d'un circuit complexe.

Dans ce cadre, notre but était d'aider le concepteur de circuits au début de la phase de conception lorsqu'il veut valider la description niveau RTL écrite à partir de la description niveau algorithmique.

Nous nous sommes focalisés sur le langage VHDL qui est un langage de description de matériel qui s'est imposé comme un standard international (Norme IEEE 1076-1993). Il est très utilisé dans l'industrie aussi bien lors de la phase de conception de circuits que lors de la phase de simulation. Du point de vue du test, une description VHDL au niveau algorithmique est similaire à un programme écrit avec un langage de programmation classique tel que C ou ADA. Cela suggère que les techniques appliquées avec succès dans le domaine du test de logiciels peuvent se révéler efficaces pour tester les conceptions de type matériel écrites avec VHDL.

Notre approche (voir figure ci-après) permet de produire ces vecteurs de test formés des stimuli (transitions des valeurs d'entrée) et des réponses attendues. Les deux descriptions (niveau algorithmique et niveau RTL) sont simulées et leurs résultats sont comparés afin de vérifier que les deux descriptions conservent le même comportement.

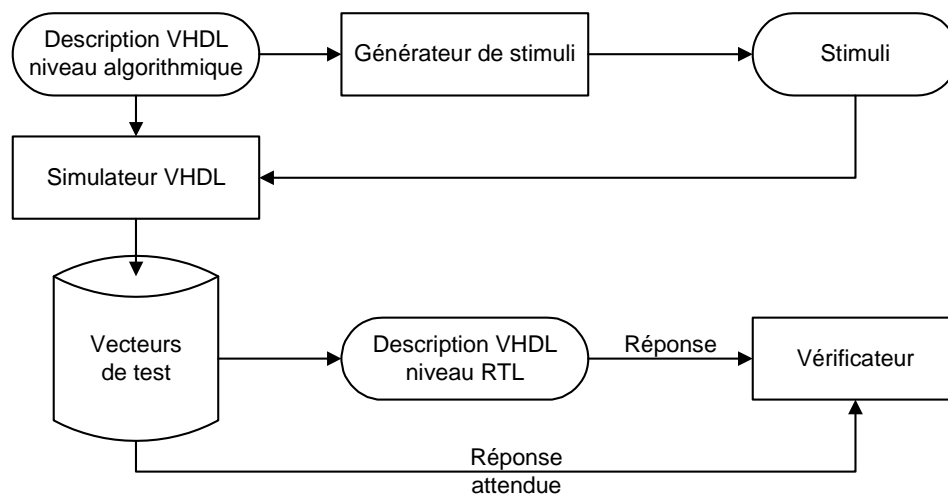


Figure 16. Aperçu de notre approche.

2. Description de l'approche

Mes travaux ont constitué en la spécification et en la réalisation d'un prototype logiciel, appelé GENESI pour GENERateur de Stimuli. Il permet de générer, à partir d'une description VHDL comportementale au niveau algorithmique, les vecteurs de test à appliquer sur une description de niveau RTL.

La conception de GENESI s'est effectuée en trois principales étapes :

1. exploration des techniques de test de logiciels ;
2. adaptation de ces techniques à VHDL ;
3. implémentation du prototype logiciel GENESI.

a) Exploration des techniques de test de logiciels

La première étape a consisté à donner une vue d'ensemble des techniques utilisées dans le domaine du test de logiciels et de choisir une technique efficace pour les descriptions de circuits digitaux écrites en langage VHDL [35].

Une étude critique de l'ensemble des techniques de test de logiciels a été présentée (test fonctionnel, test structurel). Le test structurel basé sur un critère de couverture a été décrit de manière assez détaillée car cette méthode est adaptée pour la génération de tests en utilisant le code VHDL. Cette approche est basée sur une représentation graphique du programme appelée Graphe Flot de Contrôle (GFC). Plus précisément, notre choix s'est porté sur un critère de couverture de chemins d'exécution (suite d'instructions réellement parcourues pendant l'exécution du programme) qui est le critère le plus rigoureux parmi ceux appartenant au test structurel. Or la couverture à 100% des chemins d'exécution est impossible dès qu'il y a des boucles. Cependant, le critère du test structuré, défini par McCabe [36], établit qu'un ensemble de chemins indépendants est suffisant pour tester un programme puisque tout chemin supplémentaire est combinaison linéaire de la base formée par cet ensemble. Les inconvénients de cette méthode concernent la génération de la base de chemins et la génération des jeux de test à partir de ces chemins. Pour remédier à ces problèmes nous avons présenté un algorithme de recherche « en profondeur d'abord » défini par Poole [37], qui génère une base de chemins à partir d'un GFC, et la procédure de réduction de domaine dynamique d'Offutt [38]. Cette procédure utilise une technique de satisfaction de contraintes pour extraire les jeux de test correspondant à un chemin particulier.

Ces concepts qui permettent de valider un logiciel ont été présentés dans le but de les appliquer pour détecter des erreurs de conception dans des programmes VHDL.

b) Adaptation de ces techniques à VHDL

La seconde étape a été consacrée à l'adaptation des techniques précédentes pour des programmes écrits en VHDL. Des adaptations sont nécessaires du fait de certaines caractéristiques de ce langage que l'on ne retrouve pas dans les langages de programmation traditionnels : notion de temps, interconnexion de « process » s'exécutant en parallèle, mécanisme de retard delta pour les affectations des signaux.

Nous avons ainsi présenté un GFC adapté à des programmes VHDL. Cette représentation graphique est nécessaire pour appliquer le critère du test structuré et générer la base de chemins grâce à l'algorithme de Poole. Nous avons montré en détail la correspondance entre les instructions de programmes VHDL et leur représentation graphique en tant qu'éléments du GFC. Nous avons détaillé pour chaque instruction leur structure en termes de nœud et d'arc. À partir de ce GFC, nous avons également défini la notion de chemin d'exécution ainsi que deux types de chemins particuliers : les chemins à modifier et les chemins à ordonnancer.

- i. Un chemin à modifier est un chemin de la base appartenant à un « process » fortement connecté à un autre et qui nécessite un événement sur un signal interne. On a désigné par liste des chemins solution, la liste des chemins pouvant se combiner à un chemin à modifier.
- ii. Un chemin à ordonnancer est défini comme un chemin de la base qui traverse des nœuds de décision impliquant des signaux ou des variables. À un chemin à ordonnancer correspond une liste d'ordonnancement qui permet de générer la séquence de test précédent l'exécution du chemin à ordonnancer.

Afin de résoudre les problèmes liés aux chemins à modifier et aux chemins à ordonnancer, on s'est appuyé sur deux nouvelles structures de type graphe : le « Process Model Graph » (PMG) défini par Cho et Armstrong [39] qui modélise la connectivité entre les « process », et le Graphe de Dépendance (GdD) qui modélise l'interaction (dépendance de données) entre les instructions d'affectation et les instructions de contrôle.

Concernant le problème de génération de données d'entrée (appelé stimuli dans le domaine des circuits électroniques) à partir d'un chemin, nous avons choisi de traduire notre problème vers un problème de satisfaction de contraintes (CSP en anglais pour « Constraints Solving Problem ») en nous inspirant du modèle de contrainte défini par Vemuri et Kalynaraman [40]. Nous avons vu que ces contraintes peuvent être résolues en utilisant un langage de programmation logique par contraintes (PLC) incluant un résolveur de contraintes.

À partir des structures et des concepts définis précédemment, nous avons défini une méthodologie pour la génération de vecteurs de test séparée en quatre phases distinctes :

- la production de la base de chemins qui consiste en la construction du GFC modélisant le flot de contrôle du programme VHDL et en la production de la base de chemins indépendants grâce à l'algorithme de Poole ;
- l'analyse des chemins qui permet, à partir du PMG et du GdD, la détection des chemins à modifier et des chemins à ordonnancer, générant pour chacun d'eux la liste des chemins solution et la liste d'ordonnancement. Ces listes

permettent de conserver toutes les possibilités et sont utilisables lors d'un éventuel backtracking ;

- l'extraction des vecteurs de test qui est basée sur la génération et la résolution des contraintes ;
- la production des séquences de test qui résout d'éventuels conflits (enchaînement des stimuli) et génère la séquence de test associée à chaque chemin de la base.

Nous avons présenté, pour chacune de ces quatre étapes, les différents algorithmes nécessaires à leur réalisation.

Grâce à cette approche, nous avons pu traiter le problème de génération de données d'entrée (stimuli) respectant le critère du test structuré à partir de descriptions VHDL de type comportemental.

c) Implémentation du prototype logiciel GENESI

La dernière étape de ce travail a concerné l'implémentation de ces principes en un prototype logiciel. Ce prototype logiciel appelé GENESI (pour GENERateur de Stimuli) a été réalisé pour partie en LISP, pour partie en Prolog. Il requiert l'utilisation de deux logiciels appartenant au domaine public (VHDL-1076.1 Parser/Pretty-Printer, GNU-Prolog) et d'un simulateur commercial VHDL (ModelSIM).

À partir d'un lexeur-parseur existant (VHDL-1076.1 Parser/Pretty-Printer), nous avons généré un format LISP contenant non seulement les informations sur le flot de contrôle et de données de la description VHDL mais aussi les informations concernant la déclaration et le type des objets VHDL (port, signal interne, variable, constante). Cette structure LISP ne permettant pas la séparation explicite des données et des commandes, nous avons utilisé une approche orientée objet (utilisation du langage XLISP-Plus) pour construire le Graphe de Flot de Contrôle, le Graphe de Dépendance et le Process Model Graph. L'algorithme de Poole a été implémenté pour générer une base de chemins indépendants. Cette partie du prototype logiciel a été validée sur les descriptions VHDL de type comportemental des benchmarks ITC'99.

Finalement, nous avons utilisé le système GNU-Prolog, un compilateur natif pour le langage Prolog incluant un puissant résolveur de contraintes sur les domaines finis. Nous avons montré que l'on peut traduire chaque chemin indépendant de la base en un ensemble de contraintes écrites dans le langage de PLC : clp(FD). Nous avons vu que le résolveur permet d'obtenir un ensemble de solutions pour les ports d'entrée de la description VHDL. Deux exemples concrets de description VHDL, pour lesquels GNU-Prolog a été utilisé avec succès, ont été décrits.

3. Synthèse et conclusion

Le temps d'exécution de l'algorithme principal de notre approche étant exponentiel à cause du caractère NP-complet du problème, la mise en place d'heuristiques a constitué de facto une perspective pour cette étude.

Une fois l'implémentation de GENESI terminée, nous avons mis en place différentes métriques sur les chemins qui nous permettaient d'estimer la contrôlabilité et l'observabilité des ports, signaux et variables. Nous avons utilisé ces métriques pour comparer les différents chemins solutions et faire un choix de chemin pour accélérer le temps d'exécution de nos algorithmes. La mise au point de ces métriques requière un travail de recherche long et fastidieux.

Stoppés en 2005, nous n'aborderons pas ces travaux dans la conclusion générale de ce mémoire.

VIII. Conclusion générale

Comme indiqué dans chacune des conclusions dédiées aux parties "aide à la décision" et "test de logiciel", nous ne reviendrons pas dans la conclusion générale de ce rapport sur celles-ci. Ainsi le bilan, les conclusions et les perspectives de recherches exprimés ci-après ne concernent que la problématique de la prédiction et modélisation de phénomènes spatio-temporels.

A. *Bilan et conclusions*

En préambule de ce chapitre, nous avons fait remarquer que la prévision météorologique était un problème très difficile et même indiqué qu'il s'agissait d'un problème indécidable en informatique. La quête du Graal n'étant pas atteignable, nous devons malheureusement nous contenter d'approximations.

Le problème fondamental est que les systèmes complexes tels que l'atmosphère sont extrêmement difficiles à réduire à des lois mathématiques simples et modélisés en conséquence. Les équations dans les modèles numériques ne sont donc que des approximations de la réalité, et sont souvent très sensibles aux influences extérieures et à de petits changements dans le paramétrage. Les systèmes tels que l'atmosphère ne peuvent pas être facilement simulés à l'aide d'équations, parce que leurs propriétés sont le résultat émergent [5][41][42] d'effets locaux multiples.

Dans le cadre des travaux présentés dans ce mémoire, nous nous sommes plus particulièrement intéressés à l'utilisation des réseaux de neurones artificiels appartenant aux techniques d'intelligence artificielle. Nous les avons utilisées dans le cadre de la prédiction de séries temporelles d'ensoleillement et de pics de pollution atmosphérique. Nous avons tout d'abord étudié l'impact de prétraitements statistiques des séries temporelles sur la performance de prédicteurs. Aussi, au vu du nombre important de données dont nous disposions, nous avons mis au point une méthode de sélection de caractéristiques (ou de variables) avec l'objectif d'obtenir le modèle le plus performant mais aussi le plus simple possible (principe de parcimonie). Enfin, lors de la comparaison de performance entre notre modèle et différentes méthodes de prédictions (« benchmarking » en anglais), nous avons constaté un chevauchement de celles-ci suivant l'horizon considéré. Nous avons alors étudié comment les hybrider, technique qui entre dans le cadre de la problématique plus large de la sélection de modèles. Enfin, nous avons plus récemment démarré des travaux qui concernent l'estimation de potentiel solaire, également à l'aide réseaux de neurones artificiels. Cependant, il s'agit cette fois non pas de prédire, mais d'estimer le rayonnement solaire incliné à partir du rayonnement solaire global horizontal et d'autres variables météorologiques disponibles.

L'ensemble de ces travaux ont fait l'objet de plusieurs publications dans des revues de rang A et publications dans des conférences internationales reconnues, validant ainsi les approches présentées.

Plus largement, les éléments de réponse apportés à la problématique de prédiction s'intègrent pleinement dans les stratégies de pilotage des « smart grids » ou réseaux intelligents, nouveau paradigme de la gestion des réseaux électriques. En effet, une fois qu'un « benchmark » sera établi, une hiérarchisation des différents prédicteurs permettra d'envisager la définition de stratégies de gestion intelligente de l'énergie, et donc d'algorithmes, au sein des outils de contrôle-commande de plateformes telles que MYRTE et PAGLIA ORBA. Dans cette optique, nos récents travaux à matière d'hybridation sont très encourageants.

B. Perspectives de recherche

Dès le départ de nos travaux de recherche, nous avons l'intuition qu'un couplage entre procédé expert et analyse statistique permettrait d'accroître la qualité de la prédiction. On retrouve souvent dans la littérature cette approche sous le nom de modèle de type « boîte grise » ou modélisation semi-physique. Nous souhaitons continuer dans ce sens et plusieurs perspectives s'offrent à nous. Elles concernent essentiellement la sélection de variable et l'hybridation de modèles.

En effet, concernant la problématique de la sélection de variable, notre approche a évolué vers l'utilisation de l'information mutuelle et de sa généralisation, plutôt que les coefficients de corrélation ou de corrélation partielle. Disposant d'un jeu de données très important nous souhaitons étudier le couplage de l'information mutuelle et des algorithmes stochastiques itératifs (algorithmes génétiques, recuit simulé, etc.) aussi appelés métaheuristiques. Ils ont été déjà utilisés avec succès dans la littérature dans le cadre de la sélection de variable et nous souhaitons les adapter à nos études de séries temporelles.

Concernant la problématique de l'hybridation de modèles, nous trouvons séduisante l'idée de développer des modèles partiels par classification capturant certains aspects du système. Nous envisageons de créer des scénarii de prédiction à partir de ces modèles partiels. Au moment de l'écriture de ce mémoire, nous sommes toujours au stade de l'expérimentation au sein de nos travaux sur la prédiction de pics de pollutions. A termes nous envisageons une formalisation générique de cette approche pour toutes les séries temporelles que nous étudions.

Pour la prévision de la ressource solaire, nous envisageons également d'étudier un couplage avec des données et images fournies par un imageur (« sky-imager » en anglais) afin d'affiner la prévision, ou plutôt l'erreur sur la prévision pour un horizon donné. Un « sky-imager » est un outil de mesure du ciel qui permet de compléter un

dispositif de mesures en effectuant une image du ciel, par un dispositif de miroir et d'une caméra permettant un suivi en temps réel de la nébulosité de jour.

D'autre part et plus particulièrement pour la gestion optimale des équipements de production d'énergie solaire, l'équipe EnR a souhaité investiguer l'interaction entre aérosols et rayonnement solaire. En effet, il existe actuellement très peu d'étude sur les impacts des pollutions sur le rayonnement solaire et donc leur conséquence sur la productivité des systèmes énergétiques solaires. Ces travaux constitueront la dernière étape du travail doctoral confié à M. Wani TAMAS relatif à l'étude des pics de pollution.

Enfin, nous souhaitons renforcer notre collaboration avec l'ENSAM de Paris par la poursuite des travaux engagés sur la prédiction de vitesses de vent. Un précédent travail doctoral [43] a déjà été finalisé au sein de notre laboratoire sur ce sujet et un stage de master a été effectué en 2013. Un sujet de thèse en co-tutelle avec l'ENSAM a ainsi été déposé.

IX. Références

- [1] P. Wolper, *Introduction à la calculabilité: cours et exercices corrigés*. Paris: Dunod, 2006.
- [2] K. Gödel, « Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I », *Monatshefte Für Math. Phys.*, vol. 38-38, n° 1, p. 173-198, déc. 1931.
- [3] M. Baaz, Éd., *Kurt Gödel and the foundations of mathematics: horizons of truth*. New York: Cambridge University Press, 2011.
- [4] K. Svozil, *Randomness And Undecidability In Physics*, Presumed First edition. River Edge, N.J: Wspc, 1993.
- [5] D. Orrell et P. McSharry, « System economics: Overcoming the pitfalls of forecasting models via a multidisciplinary approach », *Int. J. Forecast.*, vol. 25, n° 4, p. 734-743, oct. 2009.
- [6] M. Gardner, « Mathematical Games: The fantastic combinations of John Conway's new solitaire game "life" », *Sci. Am.*, p. 120-123, oct. 1970.
- [7] C. Calas, *Concepts et méthodes pour le météorologiste. modèles conceptuels et données disponibles Tome 1, Tome 1.* Saint-Mandé: Météo-France, 2013.
- [8] G. E. P. Box, *Time series analysis: forecasting and control*, 4th ed. Hoboken, N.J: John Wiley, 2008.
- [9] J. D. Hamilton, *Time series analysis*. Princeton, N.J: Princeton University Press, 1994.
- [10] G. Dreyfus et J.-M. Martinez, *Réseaux de neurones*. Eyrolles, 2002.
- [11] G. Dreyfus, J.-M. Martinez, M. Samuelides, et Collectif, *Apprentissage statistique*. Paris: Eyrolles, 2008.
- [12] K. Hornik, M. Stinchcombe, et H. White, « Multilayer Feedforward Networks Are Universal Approximators », *Neural Netw*, vol. 2, n° 5, p. 359-366, juill. 1989.
- [13] A. L. Erkki, E. Oja, et O. Simula, *Time Series Prediction Competition: The CATS Benchmark*. 2004.
- [14] A. Lendasse, E. Oja, O. Simula, et M. Verleysen, « Time series prediction competition: The CATS benchmark », *Neurocomputing*, vol. 70, n° 13, p. 2325-2329, 2007.
- [15] M. Diagne, M. David, P. Lauret, J. Boland, et N. Schmutz, « Review of solar irradiance forecasting methods and a proposition for small-scale insular grids », *Renew. Sustain. Energy Rev.*, vol. 27, p. 65-76, nov. 2013.
- [16] C. Voyant, « Prédiction de séries temporelles de rayonnement solaire global et de production d'énergie photovoltaïque à partir de réseaux de neurones artificiels », THESE, Université Pascal Paoli, 2011.
- [17] I. Guyon et A. Elisseeff, « An Introduction to Variable and Feature Selection », *J Mach Learn Res*, vol. 3, p. 1157-1182, mars 2003.
- [18] J. R. Rice, « The Algorithm Selection Problem », in *Advances in Computers*, vol. 15, Elsevier, 1976, p. 65-118.
- [19] V. Weber, « Caractérisation des instances difficiles de problèmes d'optimisation NP-difficiles », Université de Grenoble, 2013.

- [20] M. Graff, R. Peña, A. Medina, et H. J. Escalante, « Wind speed forecasting using a portfolio of forecasters », *Renew. Energy*, vol. 68, p. 550-559, août 2014.
- [21] C. Voyant, C. Darras, M. Muselli, C. Paoli, M.-L. Nivet, et P. Poggi, « Bayesian rules and stochastic models for high accuracy prediction of solar radiation », *Appl. Energy*, vol. 114, p. 218-226, févr. 2014.
- [22] C. Voyant, P. Haurant, M. Muselli, C. Paoli, et M.-L. Nivet, « Time series modeling and large scale global solar radiation forecasting from geostationary satellites data », *Sol. Energy*, vol. 102, p. 131-142, avr. 2014.
- [23] Y. Zhang, M. Bocquet, V. Mallet, C. Seigneur, et A. Baklanov, « Real-time air quality forecasting, part II: State of the science, current research needs, and future prospects », *Atmos. Environ.*, vol. 60, p. 656-676, déc. 2012.
- [24] Y. Zhang, M. Bocquet, V. Mallet, C. Seigneur, et A. Baklanov, « Real-time air quality forecasting, part I: History, techniques, and current status », *Atmos. Environ.*, vol. 60, p. 632-655, déc. 2012.
- [25] A. Coman, A. Ionescu, et Y. Candau, « Hourly ozone prediction for a 24-h horizon using neural networks », *Environ. Model. Softw.*, vol. 23, n° 12, p. 1407-1421, déc. 2008.
- [26] A.-L. Dutot, J. Rynkiewicz, F. E. Steiner, et J. Rude, « A 24-h forecast of ozone peaks and exceedance levels using neural classifiers and weather predictions », *Environ. Model. Softw.*, vol. 22, n° 9, p. 1261-1269, sept. 2007.
- [27] H. J. S. Fernando, M. C. Mammarella, G. Grandoni, P. Fedele, R. Di Marco, R. Dimitrova, et P. Hyde, « Forecasting PM10 in metropolitan areas: Efficacy of neural networks », *Environ. Pollut. Barking Essex 1987*, vol. 163, p. 62-67, avr. 2012.
- [28] M. Kolehmainen, H. Martikainen, T. Hiltunen, et J. Ruuskanen, « Forecasting Air Quality Parameters Using Hybrid Neural Network Modelling », *Environ. Monit. Assess.*, vol. 65, n° 1-2, p. 277-286, nov. 2000.
- [29] H.-C. Lu, J.-C. Hsieh, et T.-S. Chang, « Prediction of daily maximum ozone concentrations from meteorological conditions using a two-stage neural network », *Atmospheric Res.*, vol. 81, n° 2, p. 124-139, août 2006.
- [30] H. D. Behr, « Solar radiation on tilted south oriented surfaces: validation of transfer-models », *Sol. Energy*, vol. 61, n° 6, p. 399-413, déc. 1997.
- [31] P. Oberti, « Décision publique et recherche procédurale: illustration d'une démarche multicritère à la localisation participative d'un parc éolien en région corse », in *Journées de l'Association Française de Science Economique « Economie : aide à la décision publique »*, Université de Rennes 1, 2004, p. 34.
- [32] L. Henriët, « Systèmes d'évaluation et de classification multicritères pour l'aide à la décision: Construction de modèles et procédures d'affectation », Université Paris Dauphine - Paris IX, 2000.
- [33] P. Haurant, P. Oberti, et M. Muselli, « Multicriteria selection aiding related to photovoltaic plants on farming fields on Corsica island: A real case study using the ELECTRE outranking framework », *Energy Policy*, vol. 39, n° 2, p. 676-688, févr. 2011.

- [34] C. Paoli, P. Oberti, et M. Nivet, « Système d'information sémantique et géographique pour l'AMCD : état de l'art et construction d'une interface utilisateur », présenté à 72ème journées du groupe de travail européen « Aide multicritère à la décision », Ecole Centrale Paris, 2010.
- [35] C. Paoli, « Validation de descriptions VHDL fondée sur des techniques issues du domaine du test de logiciels », Université de Corse, 2001.
- [36] T. J. McCabe, « A Complexity Measure », *IEEE Trans. Softw. Eng.*, vol. 2, n° 4, p. 308-320, 1976.
- [37] J. Poole, « A Method to Determine a Basis Set of Paths to Perform Program Testing », Department of Commerce, NIST, Tech. rep NISTIR 5737, 1995.
- [38] A. J. Offutt, Z. Jin, et J. Pan, « The dynamic domain reduction procedure for test data generation », *Softw. Pract. Exp.*, vol. 29, n° 2, p. 167-193, févr. 1999.
- [39] C. H. Cho et J. R. Armstrong, « B-algorithm: A Behavioral Test Generation Algorithm », in *Proceedings of the 1994 International Conference on Test*, Washington, DC, USA, 1994, p. 968-979.
- [40] R. Vemuri et R. Kalyanaraman, « Generation of design verification tests from behavioral VHDL programs using path enumeration and constraint programming », *IEEE Trans. Very Large Scale Integr. VLSI Syst.*, vol. 3, n° 2, p. 201-214, juin 1995.
- [41] D. Orrell, *The Future of Everything: The Science of Prediction*. New York: Basic Books, 2008.
- [42] G. Longo et T. Paul, « Le monde et le calcul : réflexions sur calculabilité, mathématiques et physique », in *Logique & Interaction : Géométrie de la cognition*, 2009.
- [43] R. Baïle, « Analyse et modélisation multifractales de vitesses de vent. Application à la prévision de la ressource éolienne. », Université Pascal Paoli, 2010.